ELUCIDATING GENETIC DETERMINANTS OF HEART DISEASE AND LONGEVITY THROUGH GENOME-WIDE ASSOCIATION ANALYSES

by
Nathan A Bihlmeyer

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
November, 2016

Abstract

The research described in this dissertation consists of three projects. In the first, I investigate of the impact of genome-wide heterozygosity on overall longevity in humans and estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death within a given time period by 1.57%. Our data shows this to be true even if the population itself has reduced mean heterozygosity and this effect is constant between ancestry, sex, and cause of death. In the second project, I investigate of the role of coding variants in regulating QT interval, a predictor of sudden cardiac death (SCD), and identify 10 novel loci associated with QT interval, and highlight the role of 17 specific genes. Our analyses also highlight a role for the internal structure and the interconnection of myocytes in modulating QT interval duration in addition to previously implicated pathways. In the last project, I investigate of the utility of genetics in predicting implantable cardioverter-defibrillator (ICD) therapy incidence. Unfortunately, our sample size was too limited to find meaningful results. To remedy this, we are working on building a consortium to increase sample size.

Advisor: Dan E Arking, PhD

Reader: Dimitrios Avramopoulos, MD PhD

Preface

The work contained here represents 5 years of effort and learning. First and foremost, I need to acknowledge my thesis mentor, Dan E. Arking, for taking the time to teach me bioinformatics and the ability to handle large genetic datasets. When I first started in his lab, I could not run even the simplest genetic association using the simplest tool for the job, Plink. Through many hours and days and months where I abused his open-door policy, I began to turn into the research scientist he seems to think I could become. However, that was only the beginning. With endless patience and determination, Dan taught me to manage large collaborations involving weekly emails and monthly conference calls with 10s of analysts, 10s of professors, and 10s of other contributors. No one who knows me would think lightly of that kind of feat. I will be forever grateful.

I would also like to acknowledge the other two professors who let me spend time under their guidance during lab rotations, Dr. Nicholas Marsh-Armstrong and Dr. Steven Salzberg. I thank Director Dr. Dave Valle, Associate Director Dr. Kirby D. Smith, and Administrator Sandy Muscelli for accepting me into the Predoctoral Training Program in Human Genetics and Molecular Biology and guiding me through to the oral comprehensive exam and beyond. I am extremely thankful for these opportunities.

I would like to thank the members of my thesis committee: Dr. Gordon Tomaselli, Dr. Dimitrios Avramopoulos, and Dr. Eliseo Guallar. They helped me step back from my work and understand the greater meaning of what my research project entailed and then showed me how to convey that information to others.

I would like to thank all my labmates from the Arking lab, in particular: Dr. Pallav Bhatnagar, Dr. Simone Gupta, Dr. Shannon Ellis, Dr. Foram Ashar, and Anna Moes. They have been there for me every step of the way, encouraging me. They have taught me an endless amount and, as the ones who went before me, marked the way for me to follow. I could not have gotten this far without them.

I would like to thank my classmates: Amanda Price, Ana Cordova, Donghui Li, Genevieve Stein-O'Brien, Jane Welch, Jevon Cutler, Samuel Gilbert, Tyler Creamer and Xuan Pham. We fought through classes and the oral comprehensive exam together. It was a long and difficult time, without each other's help I am not sure we would have made it through.

Lastly, I would like to thank all of my collaborators, especially Jennifer A Brody and Nona Sotoodehnia. These two have been like a second lab family and mentor for me, helping me with most of my projects. Additionally, I would like to thank the analysts from around the world that have run analyses for me to contribute to my work: Aaron Isaacs, Albert Vernon Smith, Alexander Teumer, Bram P Prins, Ching-Ti

Liu, Farid Radmanesh, Gail Davies, Hao Mei, Helen R Warren, Honghuang Lin, Jeffrey Haessler, Jennifer A Smith, Jennifer E Huffman, Jessica van Setten, Jie Yao, Joanne Murabito, Jonathan Marten, Kathryn L Lunetta, Leanne M Hall, Lei Yu, Leo-Pekka Lyytikäinen, Man Li, Marten van den Berg, Martina Müller-Nurasyid, Megan L Grove, Mike Nalls, Niek Verweij, Niels Grarup, Ruifang Li-Gao, Saira Saeed Mirza, Stefan Weiss, Toshiko Tanaka, Xiuqing Guo, and Yiyi Zhang. This is the short list of people I have worked closely with. Each of them has directly done work for me to further my thesis work. For the long version of this list of my collaborators, please see the author lists on my publications. Without each and every one of my collaborators, my work would not have been possible. I am grateful.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Discussed herein are three projects I undertook broadly in the field of human genetics. The first is an investigation of the impact of genome-wide heterozygosity on overall longevity in humans. In this case we refer to the lower risk of death within a time period as “longevity.” The presence of numerous genetic variants with small effects on health and related traits raises the question of whether aggregate measures of genomic variation are associated with human health and disease. Previous studies have found increased genetic diversity associated with increased fitness across many organisms (Mitton and Grant; Alibert et al.), including humans (Roberts et al.; Coetzee et al.; Campbell et al.; Takata et al.). Two general mechanisms that act at a genome level to influence fitness have been proposed: Compensation for recessive deleterious mutations and a specific advantage of the heterozygous state over either homozygous state, called overdominance or heterozygous advantage (Charlesworth and Willis). To test for the effect of genome-wide heterozygosity on survival, we performed a meta-analysis of 17 cohorts (13 European ancestry, 4 African American ancestry) followed prospectively, with a combined sample size of 46,716 individuals, including a total of 15,234 deaths and an average age at death of 80 years old. This project is concerned with human genetics at a population level. We ask the question whether an individual's heterozygosity as compared his/her own genetic population is associated with the

individual's survival over time. Participants are followed until death. Due to about 2/3 of participants still being alive, we used the most recent available follow up data, however the individual cohort studies are still on going and these results could be updated with additional follow up data in 10 years or so. The Cox Proportional Hazards (CoxPH) model we use handles this gracefully by censoring participants at time of last follow up. Since the average age at death in this study is over the life expectancy in the areas studied, we feel this is a legitimate longevity study. It is not a goal of this study to dissect the association to locate the underlying genetic and biologic causes or identify genes and pathways that contribute to the association.

The second project is an investigation of the role of coding variants in regulating QT Interval, a predictor of SCD. QT interval, determined via a standard non-invasive electrocardiogram (EKG) by measuring the time between the beginning of the Q deflection to the termination of the T wave, is a classic measure of ventricular repolarization time. Prolonged QT interval has been linked to higher risk of SCD, with between 180,000 and 450,000 cases of SCD in the United States of America annually (Deo and Albert). Since the vast majority of SCD occurs in the absence of clinical features that would bring a victim to medical attention (Chugh et al.), identifying additional risk factors and dissecting the etiology of disease is of high importance. We conduct analyses across the Illumina

ExomeChip in population-based samples to interrogate the role of a largely unstudied class of variation on ventricular repolarization time in the population – coding variants. These variants fill in the gap between the extremely rare large-effect coding variants that result in the Mendelian long QT syndrome and short QT syndrome and the common small-effect largely non-coding variation identified through genome-wide association studies. The focus on exons and coding variants has an added benefit, in that genes can be directly implicated if LD is controlled for, whereas non-coding variation typically only implicates a region of the genome, often containing multiple genes, requiring extensive functional experiments to directly implicate a gene. We performed a meta-analysis of 241,552 variants in 17,574 genes in a sample of 95,626 individuals from 23 cohorts (comprised of 83,884 European ancestry individuals, 9,610 African Americans, 1,382 Hispanics, and 750 Asian individuals) This project focuses on a particular disease, SCD, by looking at one of its related phenotypes, QT interval, and applies human genetics to understand an individual's risk for this disease. The goal is finding individual genes that control electrophysiology and by extension affect a person's risk for SCD. This represents a narrowing of focus compared to the population-based question in the first project. Once genes are identified, they become therapeutic targets and further investigations

could lead to treatments to prevent SCD. The information we learn can also be used to assess risk of SCD in high-risk populations.

Finally, the third project is an investigation of the utility of genetics in predicting ICD therapy incidence. An accurate prediction would be useful for evaluating if a particular patient should undergo surgery to receive an ICD. ICDs are devices that can prevent SCD in patients with systolic heart failure. As stated previously, there are between 180,000 and 450,000 cases of SCD in the United States of America annually (Deo and Albert). The World Society of Arrhythmia found that there were 133,262 ICD implants in the United States in 2009 alone (Mond and Proclemer). However, ICDs administer therapy in only a minority of people in which they are implanted, about 20% in our data. This represents a significant shortcoming in the clinical selection criteria for patients at greatest risk for SCD. We wish to find out whether we can use genetics to predict if an ICD will benefit a patient. This project is concerned with applying human genetics to a clinical question. The answer to this question could directly affect the actions taken by medical doctors with regard to the manner which they treated patients. This is the narrowest focus across the three projects and could have a direct effect on people's lives.

Chapter 2: Heterozygosity and Longevity

2.1 Abstract

Background: It has been well-established, both by population genetics theory and direct observation in many organisms, that increased genetic diversity provides a survival advantage. However, given the limitations of both sample size and genome-wide metrics, this hypothesis has not been comprehensively tested in human populations. Moreover, the presence of numerous segregating small effect alleles that influence traits that directly impact health directly raises the question as to whether global measures of genomic variation are themselves associated with human health and disease.

Results: We performed a meta-analysis of 17 cohorts followed prospectively, with a combined sample size of 46,716 individuals, including a total of 15,234 deaths and an average age at death of 80 years old. We find a significant association between increased heterozygosity and survival (P=0.03). We estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death within a given time period by 1.57%.

Conclusions: This effect was consistent between European and African ancestry cohorts, men and women, and major causes of death (cancer and cardiovascular disease), demonstrating the broad positive impact of genomic diversity on human survival.

2.2 Background

With the advent of genome-wide association studies (GWAS), and more recently whole-exome and whole-genome sequencing, remarkable progress has been made in elucidating the genetics of complex traits. Numerous genetic variants, each explaining a small fraction of the phenotypic variance, have been identified ("OMIM"; Hindorff et al.). The presence in the genome of numerous segregating small effect alleles that influence health related traits raises the question of whether global measures of genomic variation are themselves associated with human health and disease. Indeed, increased fitness has been associated with the increase of genetic diversity across many organisms (Mitton and Grant; Alibert et al.), including humans (Roberts et al.; Coetzee et al.; Campbell et al.; Takata et al.), and is often referred to as positive Heterozygosity Fitness Correlations (HFCs). In particular, associations have been found between heterozygosity at the Major Histocompatibility Complex (MHC) (a.k.a. Human Leukocyte Antigen, HLA) region and general health in humans (Lie, Simmons, and Rhodes). In the case of heterozygosity in the MHC region, the cause of a positive HFC being observed is believed to be the result of increased antibody diversity conveying more robust pathogen resistance and therefore increased general health (Piertney and Oliver). However, in the case of increased

whole-genome heterozygosity, the mechanism of action is less readily apparent. Two general mechanisms that act at a genome level to influence fitness have been proposed. The first is compensation for recessive deleterious mutations (Charlesworth and Willis), whereas the second is a specific advantage of the heterozygous state over either homozygous state (overdominance/heterozygous advantage) (Charlesworth and Willis), such as that observed for the sickle cell mutation in the presence of endemic malarial disease. It has been proposed that compensation for deleterious mutations occurs at many loci and is the major mechanism at work in HFCs, with overdominance occurring at few loci but with greater effect size per occurrence (Charlesworth and Willis).

2.3 Results and Discussion

Various heterozygosity metrics have been proposed (Szulkin, Bierne, and David). The heterozygosity metric used in this study is the sum of all heterozygous loci divided by the expected state given the allele frequency under Hardy-Weinberg Equilibrium: $t = \frac{\sum 0;1}{\sum 2p(1-p)}$ where p is the frequency of the major allele in each cohort. This metric up-weights loci where the expectation of being heterozygous is low. Given

the relationship between effect size and allele frequency (Arking and Chakravarti; Hindorff, Gillanders, and Manolio), up-weighting loci with low minor allele frequencies should maximize the ability to detect a HFC in humans under a model in which the compensation for deleterious alleles is the major mechanism driving HFCs. Only Single Nucleotide Polymorphisms (SNPs) on the autosomes were considered.

To test for the effect of genome-wide heterozygosity on survival, we performed a meta-analysis of 17 cohorts (13 European ancestry, 4 African American ancestry) followed prospectively, with a combined sample size of 46,716 individuals, including a total of 15,234 deaths and an average age at death of 80 years old (Table 2.1). Within each cohort, a CoxPH was used comparing age at study entry to age at study exit (death) or most recent follow-up (alive), and included covariates known to affect survival (sex, highest education level, Body Mass Index (BMI), income level, center where DNA was collected, and the first ten principal components to adjust for population substructure). Since each cohort used a different number of SNPs (Table 2.1), the variances of the heterozygosity metrics are not the same (they are dependent on the total number of SNPs in the metric), and effect sizes from each cohort are not directly comparable. Using Stouffer's method to combine Z-scores, weighted by the number of deaths in each cohort, we find a significant association between increased heterozygosity and survival (P= 0.03). To

assess effect size, we standardized the beta estimates by multiplying them by the standard deviation of the heterozygosity metric for each cohort (Menard). This method does not completely account for the aforementioned bias; however, it is the most appropriate method to determine an interpretable effect size. Combining the standardized beta estimates using inverse variance weighting demonstrates that for every standard deviation increase in heterozygosity a person has over the population mean, they are expected to have a 1.57% decreased risk of death within a given time period (Figure 2.1). There was no evidence for heterogeneity across studies, and a direct comparison of European Ancestry to African ancestry cohorts showed no significant difference (Figure 2.2, P=0.80); thus, all downstream analyses combined European and African ancestry cohorts.

To test whether all chromosomes are contributing equally to the association between heterozygosity and survival, each study subject's heterozygosity score was recalculated using only SNPs from a given chromosome. An inverse-variance meta-analysis for each chromosome was performed across studies, followed by a meta-analysis of the chromosomal results (Figure 2.3). No significant difference was observed between effects across chromosomes (P=0.17). To test whether all major causes of death contribute equally to our genome-wide finding, death caused by cancer, death caused by CVD, and other causes of death were

each analyzed separately. A meta-analysis for each cause of death was performed as described above, followed by a test for heterogeneity and model fitting. Our results demonstrate that heterozygosity is protective for all causes of death, with no significant evidence for heterogeneity (Figure 2.4, P=0.79). To assess if heterozygosity levels impact women differently from men, meta-analyses were performed separately for each sex. Our results do not provide evidence for a differential effect of heterozygosity on survival in men vs. women (Figure 2.5, P=0.49).

2.4 Conclusions

In summary, this study provides evidence that the protective effect of increased heterozygosity seen in lower organisms functions in humans as well and may have implications for how we design future studies to identify genetic determinants of human disease and survival. We estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death within a given time period by 1.57%. Interestingly, this seems to be true even if the population itself has reduced mean heterozygosity. In future studies, limiting to heterozygosity in proximity to genes and/or regulatory elements may reveal if some regions are more sensitive to heterozygosity than others. Increasing the African ancestry sample size may increase power to see a difference between ancestry

groups. Overall the consistency we observed between European and African ancestry, males and females, and major causes of death demonstrate a broad positive impact of genomic diversity on human survival.

2.5 Tables

Table 2.1: Cohort Demographics

Included here is a descriptive breakdown of each cohort and summary statistics.

Study Name	AGES	ARIC	CHS	FHS	HealthABC
Number Samples	3209	7825	3338	4526	1448
Number Deaths	1021	1878	2220	1332	764
Percent Female	58%	53%	60.3%	54.90%	45.86%
Mean baseline Body Mass Index (BMI)	27.11	26.92	N/A	27.78	26.59
Mean followup time (Years)	7.66	19.39	13.4	9.14	10.93
Mean Age at Death (Years)	84.58	73.60	84.63	81.71	83.34
Years of baseline examinations	2002-2007	1987-1989	1989-1990	1974-2006	1997-1998
Years of DNA collection	2002-2007	1987-1998	1989-1990	1974-2006	1997-1998
Array type	Illumina Hu370CNV	Affymetrix 6.0	Illumina 370CNV	Affymetrix 500K mapping array, Affymetrix 50K supplemental gene-focused array	Illumina 1M
Genotype calling algorithm	BeadStudio	Birdseed	BeadStudio	BRLMM	GenomeStudio
Data handling and statistical tests	PLINK and R	PLINK and R	PLINK and R	PLINK, R and SAS	PLINK and R
Education Level 1 (N)	<11th grade (769)	<11th grade (1141)	<11th grade (856)	<11th grade (1315)	for those reporting grade 11 or less, also for those reporting vocation/trade school without GED (174)
Education Level 2 (N)	High school diploma, general equivalence diploma or some vocational school (1589)	High school diploma, general equivalence diploma or some vocational school (3686)	High school diploma, general equivalence diploma or some vocational school (1222)	High school diploma, general equivalence diploma or some vocational school (1788)	for high school graduates (497)
Education Level 3 (N)	1-4 years of college (487)	1-4 years of college (2558)	1-4 years of college (953)	1-4 years of college (587)	some college (777)
Education Level 4 (N)	Some graduate/professional school (367)	Some graduate/professional school (790)	Some graduate/professional school (3381)	Some graduate/ professional school (303)	N/A
Income Level 1 (N)	N/A	Under $5000 (98)	Under $5000 (92)	N/A	for those reporting less than 10k (61)
Income Level 2 (N)	N/A	$5000-$7999 (118)	$5000-$7999 (216)	N/A	for 10k to 25k (463)
Income Level 3 (N)	N/A	$8000-$11999 (261)	$8000-$11999 (340)	N/A	for >25k up to <50k (581)
Income Level 4 (N)	N/A	$12000-$15999 (389)	$12000-$15999 (489)	N/A	for 50K+ (343)
Income Level 5 (N)	N/A	$16000-$24999 (1027)	$16000-$24999 (657)	N/A	N/A
Income Level 6 (N)	N/A	$25000-$34999 (1541)	$25000-$34999 (539)	N/A	N/A
Income Level 7 (N)	N/A	$35000-$49999 (1831)	$35000-$49999 (365)	N/A	N/A
Income Level 8 (N)	N/A	Over $50000 (2578)	Over $50000 (492)	N/A	N/A
Linear Heterozygosity Whole Genome - Mean	0.9969	0.9989	0.9945	0.9973	0.9953
Linear Heterozygosity Whole Genome - SD	0.0145	0.0096	0.0116	0.0096	0.0127
Linear Heterozygosity Whole Genome - Max Number of SNPs Used	231171	371011	327843	211955	711732

HRS	INCHIANTI	LBC1921	LBC1936	MAP	ROS	Rotterdam
8617	1012	418	859	710	788	4903
1089	381	307	97	321	458	2751
58.22%	55.14%	56.22%	47.85%	72.41%	65.68%	59.36%
29.13	27.18	26.16	27.83	26.89	27.12	26.30
4.64	11.36	5.12	2.98	6.09	9.69	13.22
79.88	85.17	87.12	72.62	89.06	87.41	73.67
2006-2008	1998-2000	1999-2001	2004-2007	1997-2008	1994-2008	1990-1993
2006-2008	1998-2000	1999-2001	2004-2007	1997-2008	1994-2008	1990-1993
Illumina Human Omni2.5-4v1	Illumina 550K	Illumina 610 quad v1	Illumina 610 quad v1	Affymetrix 6.0	Affymetrix 6.0	Illumina 550K
GenomeStudio version 2011.2, Genotyping Module 1.9.4 and GenTrain version 1.0	Birdseed	Illumina GenomeStudio	Illumina GenomeStudio	Birdsuite, Broad Institute	Birdsuite, Broad Institute	The Beadstudio
PLINK, R and SAS	PLINK, R, SAS	PLINK and R	PLINK and R	PLINK and R	PLINK and R	PLINK and R
<12th grade (1117)	<6th grade (692)	<11th grade (285)	<11th grade (627)	<11th grade (50)	<11th grade (18)	<11th grade (1409)
High school diploma, general equivalence diploma (4935)	6-12 years (179)	High school diploma, general equivalence diploma or some vocational school (50)	High school diploma, general equivalence diploma or some vocational school (133)	High school diploma, general equivalence diploma or some vocational school (227)	High school diploma, general equivalence diploma or some vocational school (38)	High school diploma, general equivalence diploma or some vocational school (2555)
2yr college or 4yr college (1713)	12-16 years (98)	1-4 years of college (50)	1-4 years of college (48)	1-4 years of college (213)	1-4 years of college (42)	1-4 years of college (1361)
Master's degree or Ph.D. (852)	16+ years (43)	Some graduate/professional school (33)	Some graduate/professional school (51)	Some graduate/professional school (397)	Some graduate/professional school (711)	N/A
Under $5000 (80)	N/A	N/A	N/A	Under $5000 (15)	N/A	min-$23000 (760)
$5000-$7999 (106)	N/A	N/A	N/A	$5000-$9999 (29)	N/A	>$23000-$32000 (840)
$8000-$11999 (378)	N/A	N/A	N/A	$10000-$14999 (48)	N/A	>$32000-$40000 (869)
$12000-$15999 (481)	N/A	N/A	N/A	$15000-$19999 (78)	N/A	>$40000-$52000 (912)
$16000-$24999 (1017)	N/A	N/A	N/A	$20000-$24999 (69)	N/A	>$52000-$70000 (1092)
$25000-$34999 (1238)	N/A	N/A	N/A	$25000-$34999 (135)	N/A	>$70000 (478)
$35000-$49999 (1373)	N/A	N/A	N/A	$35000-$49999 (145)	N/A	N/A
Over $50000 (3944)	N/A	N/A	N/A	Over $50000 (222)	N/A	N/A
0.9987	0.9989	0.9991	1.0023	1.000	0.999	0.9980
0.0086	0.0064	0.0087	0.0065	0.009	0.008	0.0070
840464	384883	410009	404427	297689	297689	433844

SHIP	**ARIC - African**	**HealthABC - African**	**HRS - African**	**CHS - African**
3311	2473	995	1513	771
424	825	584	220	562
44.85%	62.38%	55.58%	63.85%	62.13%
27.19	29.77	28.49	31.14	N/A
11.23	18.58	10.10	4.52	11.95
74.00	71.98	81.89	75.90	83.98
1997-2001	1987-1989	1997-1998	2006-2008	1989-1990
1997-2001	1987-1998	1997-1998	2006-2008	1989-1990
Affymetrix 6.0	Affymetrix 6.0	Illumina 1M	Illumina Human Omni2.5-4v1	Illumina HumanOmni1-Quad_v1
Birdseed2	Birdseed	GenomeStudio	GenomeStudio version 2011.2, Genotyping Module 1.9.4 and GenTrain version 1.0	GenomeStudio
PLINK and R	PLINK and R	PLINK and R	PLINK, R and SAS	PLINK and R
<=8 years of school (1267)	<11th grade (1028)	for those reporting grade 11 or less, also for those reporting vocation/trade school without GED (437)	<12th grade (534)	<11th grade (336)
10 years of school (1475)	High school diploma, general equivalence diploma or some vocational school (736)	for high school graduates (301)	High school diploma, general equivalence diploma (730)	High school diploma, general equivalence diploma or some vocational school (210)
>10 years of school (569)	1-4 years of college (465)	some college (257)	2yr college or 4yr college (174)	1-4 years of college (142)
N/A	Some graduate/professional school (370)	N/A	Master's degree or Ph.D. (75)	Some graduate/professional school (83)
under 1375 €/month (582)	Under $5000 (340)	for those reporting less than 10k (264)	Under $5000 (61)	Under $5000 (112)
- 1875 €/month (400)	$5000-$7999 (243)	for 10k to 25k (488)	$5000-$7999 (111)	$5000-$7999 (174)
- 2375 €/month (452)	$8000-$11999 (302)	for >25k up to <50k (191)	$8000-$11999 (210)	$8000-$11999 (113)
- 2875 €/month (491)	$12000-$15999 (268)	for 50K+ (52)	$12000-$15999 (169)	$12000-$15999 (111)
- 3250 €/month (366)	$16000-$24999 (446)	N/A	$16000-$24999 (242)	$16000-$24999 (100)
- 3750 €/month (290)	$25000-$34999 (305)	N/A	$25000-$34999 (198)	$25000-$34999 (84)
- 4750 €/month (428)	$35000-$49999 (242)	N/A	$35000-$49999 (167)	$35000-$49999 (45)
over 4750 €/month (302)	Over $50000 (176)	N/A	Over $50000 (355)	Over $50000 (32)
0.9991	0.9986	0.9968	0.9984	0.9958
0.0089	0.0156	0.0202	0.0120	0.0278
389567	386600	757443	919711	654600

2.6 Figures

Figure 2.1: Heterozygosity Meta-Analysis by Study

1.57% decreased risk of death for every standard deviation increase in heterozygosity. This is determined using an inverse variance weighted fixed effect model. Significance of P=0.03 is determined using Stouffer's method to combine Z-scores due to bias in inverse variance weighted fixed effect model. There are 46,716 individuals, including a total of 15,234 deaths and an average age at death of 80 years old. EA = European Ancestry; AA = African Ancestry; AGES = Age, Gene/Environment Susceptibility cohort; ARIC = Atherosclerosis Risk In Communities cohort; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; HealthABC = HealthABC cohort; HRS = Health and Retirement Study; INCHINTI = InCHIANTI cohort; LBC1921 = 1921 Lothian Birth Cohort; LBC1936 = 1936 Lothian Birth Cohort; MAP = Rush Memory and Aging Project cohort; ROS = Religious Orders Study; Rotterdam = Rotterdam Study; SHIP = Study of Health In Pomerania cohort; SE = Standard Error; HR = Hazard Ratio; CI = Confidence Interval; W = Weight; N = Number

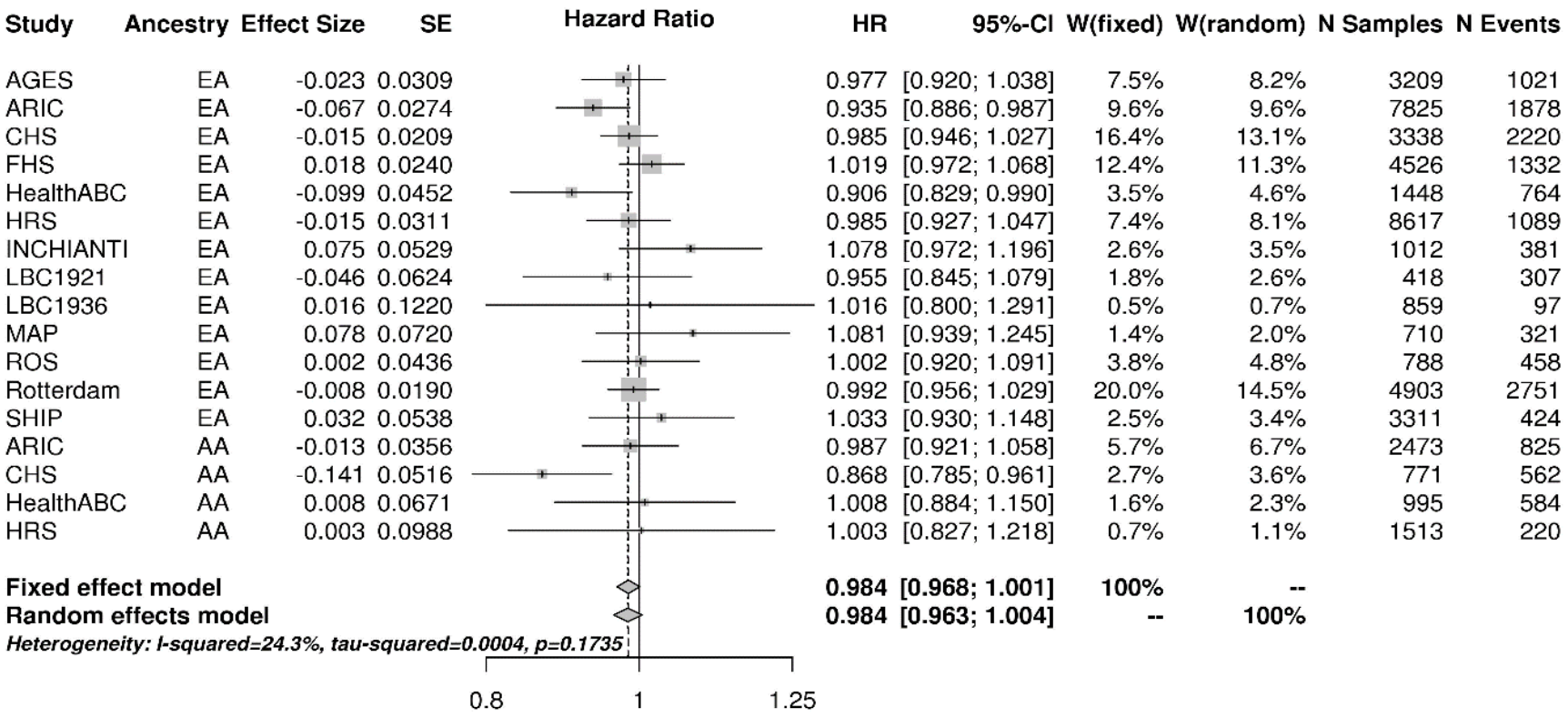

Study Ancestry Effect Size SE Hazard Ratio HR 95%-CI W(fixed) W(random) N Samples N Events
AGES EA -0.023 0.0309 0.977 [0.920; 1.038] 7.5% 8.2% 3209 1021
ARIC EA -0.067 0.0274 0.935 [0.886; 0.987] 9.6% 9.6% 7825 1878
CHS EA -0.015 0.0209 0.985 [0.946; 1.027] 16.4% 13.1% 3338 2220
FHS EA 0.018 0.0240 1.019 [0.972; 1.068] 12.4% 11.3% 4526 1332
HealthABC EA -0.099 0.0452 0.906 [0.829; 0.990] 3.5% 4.6% 1448 764
HRS EA -0.015 0.0311 0.985 [0.927; 1.047] 7.4% 8.1% 8617 1089
INCHIANTI EA 0.075 0.0529 1.078 [0.972; 1.196] 2.6% 3.5% 1012 381
LBC1921 EA -0.046 0.0624 0.955 [0.845; 1.079] 1.8% 2.6% 418 307
LBC1936 EA 0.016 0.1220 1.016 [0.800; 1.291] 0.5% 0.7% 859 97
MAP EA 0.078 0.0720 1.081 [0.939; 1.245] 1.4% 2.0% 710 321
ROS EA 0.002 0.0436 1.002 [0.920; 1.091] 3.8% 4.8% 788 458
Rotterdam EA -0.008 0.0190 0.992 [0.956; 1.029] 20.0% 14.5% 4903 2751
SHIP EA 0.032 0.0538 1.033 [0.930; 1.148] 2.5% 3.4% 3311 424
ARIC AA -0.013 0.0356 0.987 [0.921; 1.058] 5.7% 6.7% 2473 825
CHS AA -0.141 0.0516 0.868 [0.785; 0.961] 2.7% 3.6% 771 562
HealthABC AA 0.008 0.0671 1.008 [0.884; 1.150] 1.6% 2.3% 995 584
HRS AA 0.003 0.0988 1.003 [0.827; 1.218] 0.7% 1.1% 1513 220
Fixed effect model 0.984 [0.968; 1.001] 100% --
Random effects model 0.984 [0.963; 1.004] -- 100%
Heterogeneity: I-squared=24.3%, tau-squared=0.0004, p=0.1735
0.8
1
1.25

Figure 2.2: Ancestry Meta-Analysis

Direct comparison of European Ancestry to African ancestry cohorts showed no significant difference (P=0.80). Figure is formatted the same as Figure 2.1.

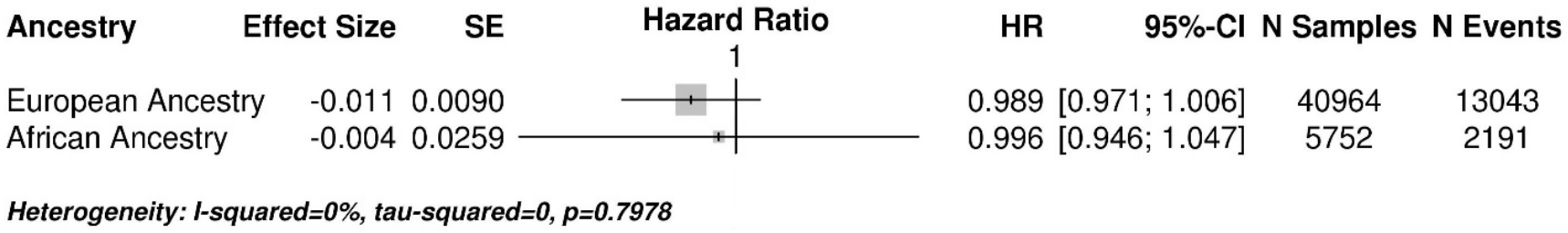
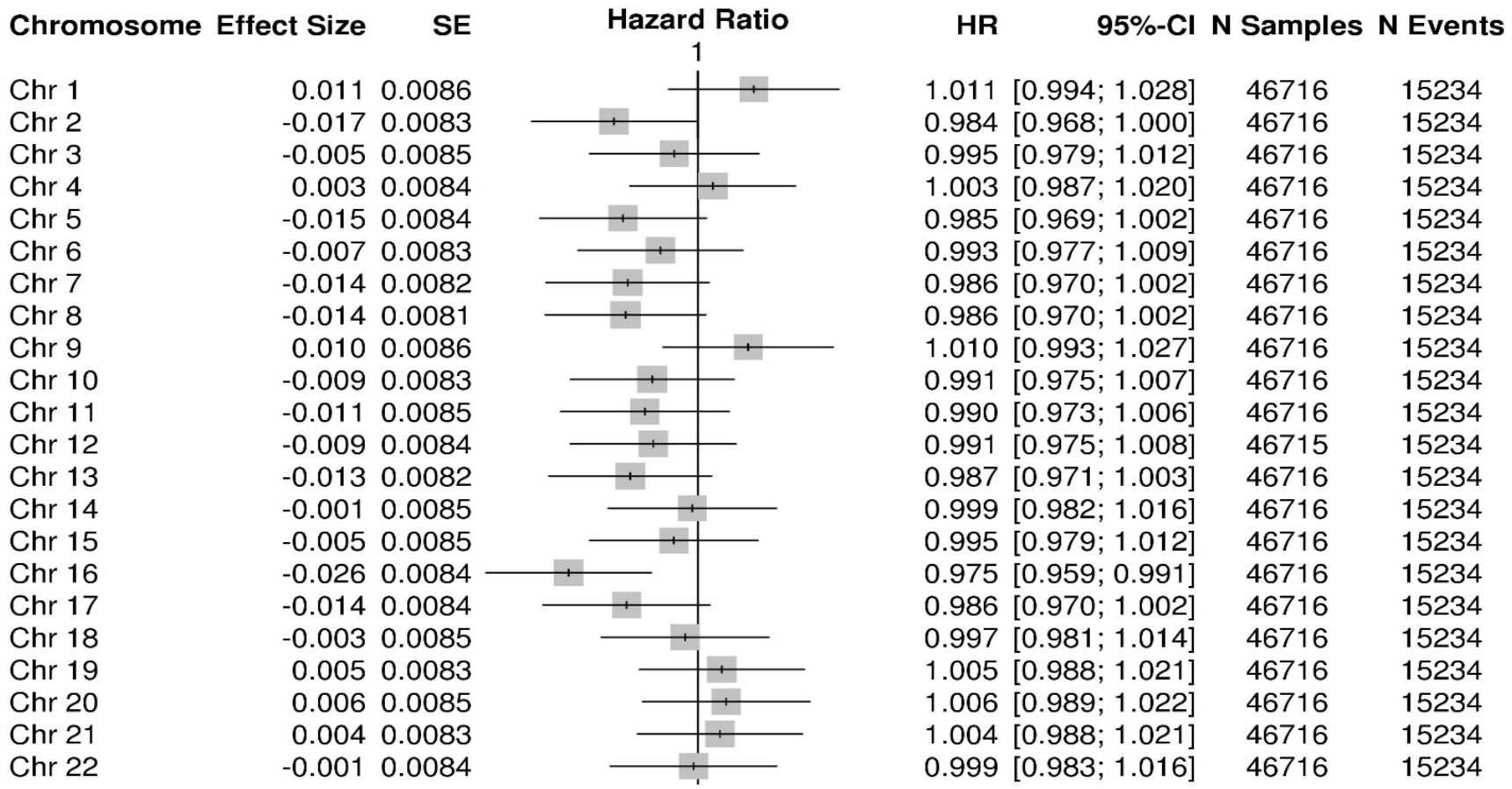

Ancestry
Effect Size
SE
Hazard Ratio
1
HR
95%-CI
N Samples
N Events
European Ancestry
-0.011
0.0090
0.989
[0.971; 1.006]
40964
13043
African Ancestry
-0.004
0.0259
0.996
[0.946; 1.047]
5752
2191
Heterogeneity: I-squared=0%, tau-squared=0, p=0.7978

Figure 2.3: Chromosome Meta-Analysis

A meta-analysis for each chromosome was performed across studies. No significant difference was observed between effects across chromosomes (P=0.17). Figure is formatted the same as Figure 2.1.



Heterogeneity: I-squared=22.4%, tau-squared<0.0001, p=0.1689

Figure 2.4: Causes of Death Meta-Analysis

A meta-analysis for each cause of death was performed. Our results show no significant evidence for heterogeneity (P=0.79). Figure is formatted the same as Figure 2.1.

Cause of Death	Effect Size	SE	Hazard Ratio	HR	95%-CI	N Samples	N Events
			1				
Cancer Death	-0.021	0.0202		0.979	[0.941; 1.019]	27502	2545
CVD Death	-0.022	0.0164		0.978	[0.947; 1.010]	27502	3859
Other Death	-0.038	0.0202		0.963	[0.925; 1.001]	24164	3193

Heterogeneity: I-squared=0%, tau-squared=0, p=0.7904

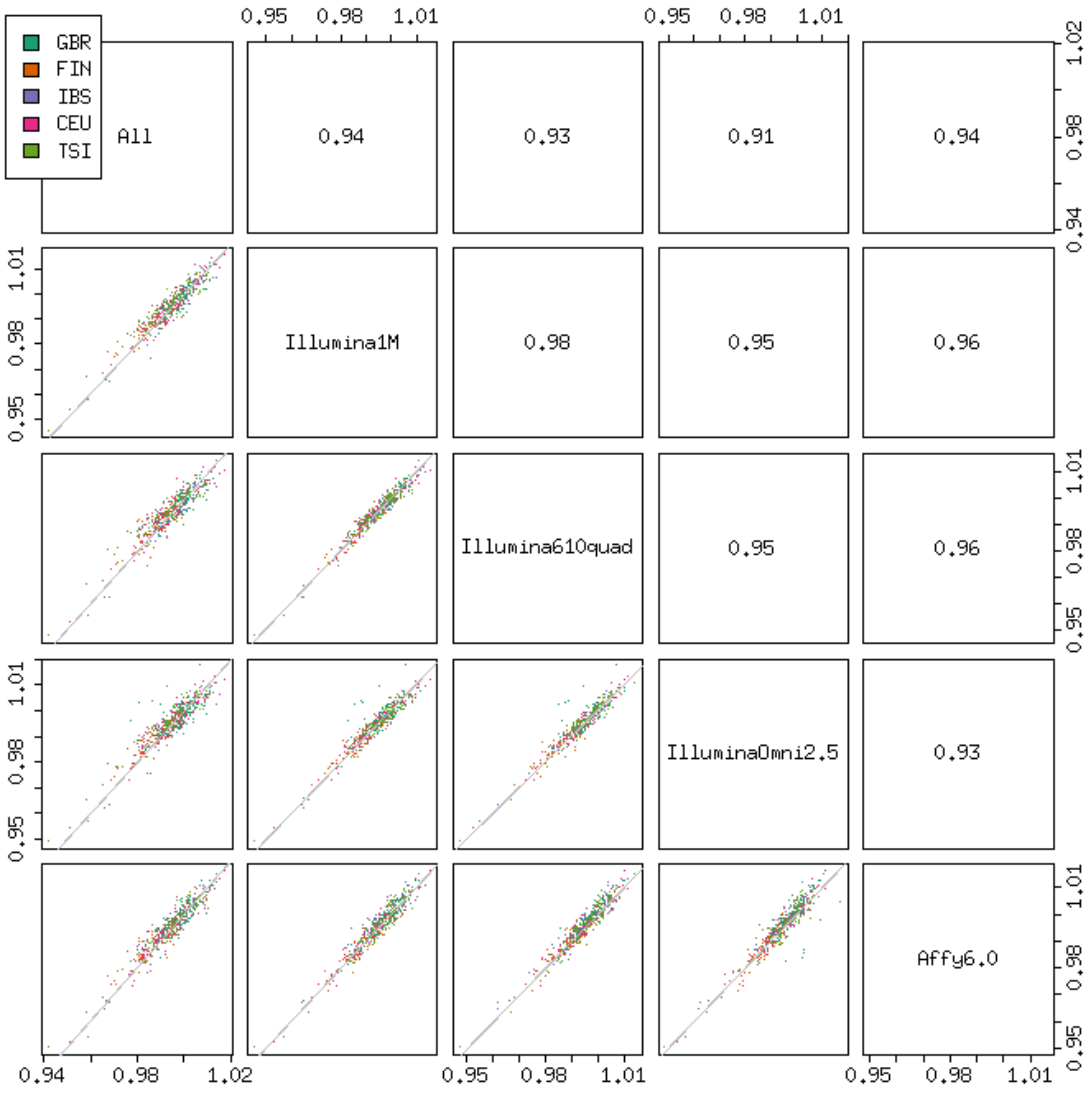
Figure 2.5: Sex Meta-Analysis

A meta-analysis was performed separately for each sex. Our results do not provide evidence for a differential effect of heterozygosity on survival in men vs. women (P=0.49). Figure is formatted the same as Figure 2.1.

Heterogeneity: I-squared=0%, tau-squared=0, p=0.4941

Figure 2.6: Heterozygosity Determined Using Various SNP Lists

Also included as a supplement is a figure showing the relationship between heterozygosity metrics determined using different SNP lists. The dataset used was genome wide SNP data from sequencing of 503 individuals with European ancestry from 1000G phase 3 release. The SNP lists used were: 1) all SNPs 2) SNPs on the Illumina 1M 3) SNPs on the Illumina 610quad 4) SNPs on the Illumina Omni2.5 and 5) SNPs on the Affymetrix 6.0. This is to determine if SNP selection on the arrays biases the heterozygosity metric. We see high correlation and no systematic bias.


GBR
FIN
IBS
CEU
TSI
All
0.94
0.93
0.91
0.94
Illumina1M
0.98
0.95
0.96
Illumina610quad
0.95
0.96
IlluminaOmni2.5
0.93
Affy6.0
0.95 0.98 1.01
0.94 0.98 1.02

2.7 Methods – Heterozygosity

Self-described Caucasian ("white", "Caucasian") and African ancestry ("black", "African American") individuals were included after excluding first and second-degree relatives and genetic outliers. Genetic outliers were defined by merging genotyping data with HapMap3 data, and calculating the Euclidean distance from a combined reference HapMap3 population (Caucasian=CEU+TSI, African ancestry=ASW+YRI+MKK+LWK) cluster centroid in the first 3 PC space weighted by explained variance. Specifically, the standard deviation of Euclidean distance was determined for each HapMap reference group, and any sample greater than ten standard deviations away from centroid were defined as genetic outliers and excluded.

Directly genotyped SNPs were used for all analyses. Imputed SNPs were not used to avoid issues with genotype accuracy and bias towards the reference panel. SNP exclusion criteria included: monomorphic in the dataset, non-unique mapping to Hg19, SNPs which are no longer in the company provided annotation file for the SNP array, >0.5% missing data, MAF ≤ 10%, HWE p-value ≥ 0.001, and non-autosomal SNPs. The heterozygosity metric is the sum of all heterozygous loci divided by the

expected state given the allele frequency under Hardy-Weinberg

Equilibrium: $t = \frac{\sum 0;1}{\sum 2p(1-p)}$ where p is the frequency of the major allele.

Separate association analyses were run for Caucasian and African ancestry samples from each cohort. The CoxPH model included covariates for Body Mass Index (BMI) at first visit and first ten principal components, and the 'strata' function for sex, education level (defined as 1. ≤11th grade, 2. high school diploma, general equivalence diploma or some vocational school, 3. 1–4 years of college, 4. Some graduate/professional school, and Missing), income level (defined by cohorts), and center of DNA collection within cohorts. The CoxPH model was set up so that the outcome was age at study entry, age at study exit, and a binary variable coding state of death (1: Dead, 0: Alive). Age is measured in units of years, but is accurate to the nearest day.

For the meta-analysis, significance was determined by Stouffer's method (Stouffer Samuel et al.) calculated as a two-sided test by incorporating Z-scores derived from two-sided tests performed in each cohort. We standardized the beta estimates by multiplying them by the standard deviation of the heterozygosity metric for each cohort, to account for the fact that the effect size is proportional to the variance in the heterozygosity metric. The variance heterozygosity metric in turn is proportional to the inverse of the square root of the number of SNPs used

to determine the heterozygosity metric. Because most cohorts used different genotyping arrays, a large bias is introduced into the meta-analysis. Stouffer's method completely removes this bias; however, cannot estimate a combined effect size, only the overall significance. To get an estimate of the combined effect size (recognizing that the P-value and associated confidence intervals will be inflated), we used inverse variance weighting of the standardized cohort effect sizes, which partially corrects the bias and allows for the combined effect size to be estimated.

2.8 Methods – Cohorts

ARIC

The Atherosclerosis Risk In Communities (ARIC) study was established in 1986 as a prospective study of 15,792 individuals, 45-65 years of age, from 4 different US communities (Jackson, Mississippi; Forsyth County, North Carolina; Washington County, Maryland; suburbs of Minneapolis, Minnesota). The first visit was carried out in 1987-89, with four subsequent in-person visits and annual telephone interviews after initial visit. Mortality was tracked via telephone follow-ups, hospitalization records, state records, and the National Death Index. Each visit consisted of data collected about electrocardiographic measures and cardiovascular outcomes. Cause of death was determined using cause of

death on the death certificate (ICD code). Only samples with a self-reported race of white or black were included in this analysis ("The Atherosclerosis Risk in Communities (ARIC) Study").

AGES

The Age, Gene/Environment Susceptibility (AGES Reykjavik) Study (Tamara B Harris et al., "Age, Gene/Environment Susceptibility-Reykjavik Study") was initiated to examine genetic susceptibility and gene/environment interaction as these contribute to phenotypes common in old age. The Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 people attended, resulting in 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages. One group was designated a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGES-Reykjavik study re-examined 5764 survivors of the original cohort who had participated before in the Reykjavik Study. Successful genotyping was available for 3219 AGES participants who were eligible for this study. The AGES Reykjavik Study

GWAS was approved by the National Bioethics Committee and the Data Protection Authority.

CHS

The Cardiovascular Health Study (CHS) (Fried et al.) is a population-based cohort study of risk factors for CHD and stroke in adults ≥65 years conducted across four field centers. The original predominantly Caucasian cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons were enrolled for a total sample of 5,888. DNA was extracted from blood samples drawn on all participants at their baseline examination in 1989-90. In 2007-2008, genotyping on the European ancestry individuals was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system on the CHS participants who were free of CVD at baseline, consented to genetic testing, and had DNA available for genotyping. In 2010, genotyping on the African-American CHS participants who consented to genetic testing, and had DNA available for genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina HumanOmni1-Quad_v1 BeadChip system.

FHS

The Framingham Heart Study (DAWBER, MEADORS, and MOORE; Feinleib et al.; Kannel et al.; Splansky et al.) is a community-based family study initiated to study determinants of cardiovascular and other chronic diseases. The study started in 1948 with the enrollment of 5,209 men and women who have been examined every two years since study inception. In 1971, 5,124 children of the original cohort and spouses of the children were enrolled in the Offspring cohort and have been examined every 4 to 8 years. Beginning in 2002, 4,095 adult grandchildren of the original cohort (children of the Offspring cohort) were enrolled into the Third Generation cohort (Gen 3) and have completed two examinations. Blood samples for DNA collection were obtained on the surviving original cohort and Offspring members in the 1990s and on the Gen 3 participants on study entry. Participants from all three cohorts are under continuous surveillance for death. For the purposes of this genetic heterozygosity investigation, Original cohort and Offspring participants were followed from the time of DNA draw until death or end of follow –up in 2009. There were 4525 participants (2485 women) included in the study sample and 1332 deaths (664 deaths in women, mean age at death 81.7 years) occurred during follow-up (mean follow –up time 9.1 years). CoxPH models were used to determine the

linear effect of heterozygosity on survival. A robust variance estimate clustering on family was used to account for family correlation. Additional models examined the association of quintiles of heterozygosity with survival. As described in the analysis section, the models were adjusted for body mass index at baseline and principal components 1-10, stratified by sex and education category. In secondary analyses, sex specific models and by-chromosome heterozygosity were examined. Income at baseline was not available and therefore not included in the model.

HealthABC

HealthABC (Tamara B. Harris et al.). Genomic DNA was extracted from buffy coat collected using PUREGENE® DNA Purification Kit during the baseline exam. In 2009, genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1M-Duo BeadChip system. Samples were excluded from the dataset for the reasons of sample failure (call rate < 95%), genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. African American and European ancestry were confirmed using principal components analyses with HapMap 3 populations as references. Genotyping was successful in 2,802 individuals (1663 European

ancestry and 1139 African Americans). Genotypes were available on 914263 high quality SNPs.

HRS

The Health and Retirement Study (HRS) (Juster and Suzman) is a longitudinal survey of a representative sample of Americans over the age of 50. The current sample is over 26,000 persons in 17,000 households. The study interviews respondents every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006 and 2008.

InCHIANTI

The InCHIANTI study (Ferrucci et al.) is a population-based epidemiological study aimed at evaluating the factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy. The details of the study have been previously reported. Briefly, 1616 residents were selected from the population registry of Greve in Chianti (a rural area: 11,709 residents with 19.3% of the population greater than 65 years of age), and Bagno a Ripoli (Antella village near

Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n=1453), and the subjects ranged between 21-102 years of age. Illumina Infinium HumanHap 550K SNP arrays were used for genotyping. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD).

LBC

Lothian Birth Cohorts 1921 and 1936 (LBC1921, LBC1936) (Deary, Whiteman, et al.; Deary, Gow, Taylor, et al.; Deary, Gow, Pattie, et al.) The Lothian Birth Cohort 1936 (LBC1936) consists of 1,091 relatively healthy individuals assessed on cognitive and medical traits at about 70 years of age. They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland in older age. At recruitment the sample of 548 men and 543 women had a mean age 69.6 years (SD = 0.8). Two further waves were carried out at 73 (n = 866 (448 males, 418 females) and 76 years of age (in progress). A full description of participant recruitment and testing can be found elsewhere. At recruitment the LBC1921 cohort consisted of 550 relatively healthy individuals, 316 females and 234 males, assessed on cognitive and medical traits at about 79 years of age. Three further testing waves were completed at 83 (n = 321 (145 males,

176 females), 87 (n = 235 (109 males, 126 females) and 90 (n = 129 (59 males, 70 females) years of age. They were all born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) in Scotland in older age. When tested, the sample had a mean age of 79.1 years (SD = 0.6). A full description of participant recruitment and testing can be found elsewhere. Ethics permission for the study was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LBC1936: LREC/2003/2/29 and LBC1921: LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent.

MAP/ROS

Religious Orders Study (ROS) (Bennett, Schneider, Arvanitakis, et al.) The ROS, started in 1994, is a longitudinal, clinical-pathologic cohort study of common chronic conditions of aging. The study enrolls Catholic priests, nuns, and brothers from about 40 groups in 12 states of the United States. Since January 1994, over 1,100 participants completed their baseline evaluation. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Participants were free of known dementia at

enrollment, agreed to annual clinical evaluations, and signed both an informed consent and an Anatomic Gift Act form donating their brains at time of death. DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute.

Rush Memory and Aging Project (MAP) (Bennett, Schneider, Buchman, et al.)

The MAP, started in 1997, is a longitudinal, clinical-pathologic cohort study of common chronic conditions of aging. The study enrolled older men and women from assisted living facilities in the Chicago area with no evidence on dementia at baseline. Since October 1997, over 1,500 participants completed their baseline evaluation. The study was approved by the institutional review board of Rush University Medical Center. The follow-up rate of survivors exceeds 90%. Similar to the ROS, participants agreed to annual clinical evaluations and signed both an informed consent and an Anatomic Gift Act form donating their brains, spinal cords, and selected nerves and muscles to Rush investigators at the time of death. DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute's Center for Genotyping and the Translational Genomics Research Institute.

Rotterdam

The Rotterdam Study (Hofman, Breteler, et al.) is a prospective cohort study ongoing since 1990 in the city of Rotterdam in The Netherlands. The study targets cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, dermatological, oncological, and respiratory diseases. As of 2008, 14,926 subjects aged 45 years or over comprise the Rotterdam Study cohort. The findings of the Rotterdam Study have been presented in over a 1,000 research articles and reports (www.erasmus-epidemiology.nl/rotterdamstudy).

SHIP

The Study of Health in Pomerania (SHIP) (John et al.; Völzke et al.) is a cross-sectional survey in West Pomerania, the north-east area of Germany. A sample from the population aged 20 to 79 years was drawn from population registries. First, the three cities of the region (with 17,076 to 65,977 inhabitants) and the 12 towns (with 1,516 to 3,044 inhabitants) were selected, and then 17 out of 97 smaller towns (with less than 1,500 inhabitants), were drawn at random. Second, from each of the selected communities, subjects were drawn at random, proportional to the population size of each community and stratified by age and gender. Only individuals with German citizenship and main

residency in the study area were included. Finally, 7,008 subjects were sampled, with 292 persons of each gender in each of the twelve five-year age strata. In order to minimize drop-outs by migration or death, subjects were selected in two waves. The net sample (without migrated or deceased persons) comprised 6,267 eligible subjects. Selected persons received a maximum of three written invitations. In case of non-response, letters were followed by a phone call or by home visits if contact by phone was not possible. The SHIP population finally comprised 4,308 participants (corresponding to a final response of 68.8%). Genotyping: The SHIP samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. Hybridisation of genomic DNA was done in accordance with the manufacturer's standard recommendations. The genetic data analysis workflow was created using the Software InforSense. Genetic data were stored using the database Caché (InterSystems). Genotypes were determined using the Birdseed2 clustering algorithm. For quality control purposes, several control samples where added. On the chip level, only subjects with a genotyping rate on QC probesets (QC callrate) of at least 86% were included. Finally, all arrays had a sample callrate > 92%. The overall genotyping efficiency of the GWA was 98.55%. Duplicate samples as estimated by IBD and individuals with mismatch between reported and genotyped gender were excluded from analysis.

2.9 Ethics Statements

Institutional Review Board approvals were obtained by each participating ARIC study center (the Universities of NC, MS, MN, and John Hopkins University) and the coordinating center (University of NC), and the research was conducted in accordance with the principles described in the Helsinki Declaration. All subjects in the ARIC study gave informed consent. For more information see dbGaP Study Accession: phs000280.v2.p1. JHSPH IRB number H.34.99.07.02.A1. Manuscript proposal number MS1964.

HealthABC Human subjects protocol UCSF IRB is H5254-12688-11.

CHS was approved by institutional review committees at each site, the subjects gave informed consent, and those included in the present analysis consented to the use of their genetic information for the study of cardiovascular disease. It is the position of the UW IRB that these studies of de-identified data, with no patient contact, do not constitute human subjects research. Therefore we have neither an approval number, nor an exemption.

IRB permission to conduct genetics-related work in the Health and Retirement Study (HRS) is granted under the project title, "Expanding a National Resource for Genetic Research in Behavioral & Health Science"

(HUM00063444). The IRB that approved this project is the Health Sciences and Behavioral Sciences Institutional Review Board at the University of Michigan. No manuscript proposal is required for use of HRS data.

Inchianti ethics review statement: The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD).

The Religious Orders Study (ORA# 91020181) and the Rush Memory and Aging Project (ORA# 86121802) were approved by the Institutional Review Board of Rush University Medical Center. Written informed consent was obtained from all the participants.

The SHIP study followed the recommendations of the Declaration of Helsinki. The study protocol of SHIP was approved by the medical ethics committee of the University of Greifswald. Written informed consent was obtained from each of the study participants. The SHIP study is described in PMID: 20167617

The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. A written informed consent was obtained from all participants.

The Boston University Medical Campus Institutional Review Board approved the FHS genome-wide genotyping (protocol number H-226671) and genetic investigation of aging and longevity phenotypes (protocol number H-24912).

The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, (VSN: 00-063) and the Data Protection Authority. The researchers are indebted to the participants for their willingness to participate in the study.

Ethics permission for the LBC studies was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LBC1936: LREC/2003/2/29 and LBC1921: LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent.

Chapter 3: ExomeChip and QT interval

3.1 Abstract

QT interval, measured through a standard EKG, captures the time it takes for the ventricles in the heart to depolarize and repolarize. JT interval, a measure of ventricular repolarization time alone, can be mathematically derived by subtracting the QRS interval, a measure of ventricular depolarization time, from the QT interval. Prolonged QT interval has been linked to higher risk of SCD. We performed an exome-wide analysis for both QT and JT intervals, including both common and rare variants from the Illumina Infinium HumanExome BeadChip using single variant statistical models.

We performed a meta-analysis of 241,552 variants in 17,574 genes in a sample of 95,626 individuals from 23 cohorts (comprised of 83,884 European ancestry individuals, 9,610 African Americans, 1,382 Hispanics, and 750 Asian individuals) and identified 10 loci that modulate QT interval and/or JT interval that have not been previously reported in the literature. This brings the total number of ventricular repolarization time associated loci to 45. Additionally, our approach of using coding variants has highlighted the role of 17 specific genes in ventricular repolarization time regulation, 7 of which are in novel loci.

3.2 Author Summary

We investigate the genetic effect coding variants have on ventricular repolarization time and in doing so implicate 17 specific genes as regulating repolarization time. Our analyses show a role for the internal structure of myocytes and interconnection of myocytes in modulating QT interval duration, adding to previous known roles of potassium ion regulation, sodium ion regulation, calcium ion regulation, and autonomic control. We anticipate these discoveries will open new paths to the goal of making novel remedies for the prevention of lethal ventricular arrhythmias and SCD.

3.3 Introduction

QT interval, determined via a standard non-invasive EKG, is a classic measure of ventricular repolarization time. Prolonged QT interval has been linked to higher risk of SCD, with between 180,000 and 450,000 cases of SCD in the United States of America annually (Deo and Albert). Since the vast majority of SCD occurs in the absence of clinical features that would bring a victim to medical attention (Chugh et al.), identifying additional risk factors and dissecting the etiology of disease is of high importance.

Heritability estimates of QT interval are between 30 and 40%, indicating that genetic variants play a large role in modulating QT

interval in the general population (Newton-Cheh, Larson, et al. 3). In addition, Mendelian syndromes of QT interval (Long QT Syndrome [LQTS] and Short QT Syndrome [SQTS]) occur in ~1:2000 individuals, and are caused by variants in ion channels or their interacting proteins (Schwartz, Crotti, and Insolia).

Previous candidate gene and genome-wide association studies (GWAS), largely screening common non-coding variants, have identified 35 loci containing variants of small effect (Pfeufer et al.; Newton-Cheh, Eijgelsheim, et al.; Arking, Pfeufer, et al. 1; Nolte et al.; Holm et al.; Noseworthy et al.; Kim et al.; Arking, Pulit, et al.), with the largest study, QTIGC (Arking, Pulit, et al.), including a discovery population of 76,061 European ancestry individuals. Together, familial and population-based studies of QT interval have highlighted prominent roles for potassium ion regulation via observed associated variation in *KCNQ1* (*LQT1*, *SQT2*), *KCNH2* (*LQT2*, *SQT1*), *KCNJ2* (*LQT7*, *SQT3*), and *KCNE1* (*LQT5*), sodium ion regulation via observed associated variation in *SCN5A* (*LQT3*), *CAV3* (*LQT9*), and *SNTA1* (*LQT10*), and calcium ion regulation via observed associated variation in *CACNA1C* (*LQT8*), *ATP2A2*, *PLN*, *PRKCA*, *SRL*, and *SLC8A1*.

In this study, we conduct exome-wide analyses in population-based samples to interrogate the role of a largely unstudied class of variation on ventricular repolarization time in the population – coding

variants. These variants fill in the gap between the extremely rare large-effect coding variants that result in the Mendelian L/SQT syndromes and the common small-effect largely non-coding variation identified through GWAS. The focus on exons and coding variants has an added benefit, in that genes are directly implicated, whereas non-coding variation typically only implicates a region of the genome, often containing multiple genes, requiring extensive functional experiments to directly implicate a gene. Despite extensive bioinformatics analyses, QTIGC only narrowed down the list of candidate genes to 48 in 25 loci, with no clear candidates in the remaining 10 loci. Our study, with its focus on coding variation, facilitates a direct link between gene and phenotype.

We performed a meta-analysis of 23 cohorts, including 95,626 multi-ethic individuals comprised of 83,884 European ancestry individuals, 9,610 African Americans, 1,382 Hispanics, and 750 Asian individuals (Table 3.3). Each individual was genotyped for 241,552 variants in 17,574 genes using the Illumina Infinium HumanExome BeadChip (ExomeChip). These variants were chosen by evaluating approximately 12,000 exome sequences for coding variants that appeared in at least two individuals along with non-coding variants of known importance from previous GWAS and variants tiling across the genome. Organizing of the sequencing, genotyping, and phenotyping was done as a part of the Cohorts for Heart and Aging Research in Genomic

Epidemiology (CHARGE) Consortium (Psaty et al.) EKG and ExomeChip working groups.

3.4 Results

6 Novel Loci Associated with QT Interval

Single nucleotide variants (SNVs) in 25 loci were associated with QT interval at exome-wide significance. Of these, 19 loci were previously identified and 6 loci were novel. All 41 QT interval known and novel associated loci are shown in Table 3.1, with the most significant SNV from the current analyses shown. The 6 novel loci are listed by nearest gene, and for four of these loci (*PM20D1*, *SLC4A3*, *CASR*, *NRAP*), the top hit is a nonsynonymous variant, providing direct evidence that the gene is involved in modulating QT interval.

Utilization of Coding Variants to Implicate Genes

For genes in the 35 previously identified loci for which the top SNV in our study is putative functional coding variant (14 loci), we wanted to determine whether the associated variant implicates a causal gene, or alternatively, is associated due to linkage disequilibrium with a more strongly associated non-coding variant, and thus would not provide evidence for the gene being involved in modulating QT interval. We

therefore determined the correlation between the top SNV in the current study, and the previously reported most significant SNV from QTIGC in ARIC European ancestry individuals (Table 3.4). We saw moderate-to-high LD (r^2>0.5) in 5 of the 14 loci. We then performed conditional analyses by including both variants in the same regression model in the same ARIC Europeans dataset (Table 3.4). Conditional analyses demonstrate that the *TTN* coding variant explains the common variant association initially reported by QTIGC. For four additional loci, the nonsynonymous coding variant is independent of the QTIGC signal, supporting a specific gene within the locus: *SP3*, *SPATS2L*, *MATN2* at the "*LAPTM4B*" locus, and *DCAF13* at the "*AZ1N1*" locus. Although the coding variants in those last four genes are not exome-wide significant. (Locus names are taken from previously published studies on QT interval; namely QTIGC (Arking, Pulit, et al.).)

As we have noted previously (Arking, Pulit, et al.), several loci contain multiple independent genetic effects, including some loci harboring multiple exome-wide significant coding variants (Table 3.5). Thus, even if not the top hit at a locus, putative functional SNVs can still implicate a specific gene at a locus. We used the Gene-Wide Significance (GWiS) (Huang et al.) software to determine the number of independent effects in all loci along with a SNV that best represents each independent effect (Table 3.6). The *SCN5A-SCN10A* locus is a particularly illustrative

example of the utility of this approach. Coding variants in *DLEC1*, *SCN5A*, and *SCN10A* are each exome-wide significant (Table 3.2). However, after utilizing GWiS, the signal coming from the coding variants in *DLEC1* and *SCN5A* is explained by non-coding variants and only the *SCN10A* coding variant signal remains. Included in Table 3.1 is each gene that GWiS has found an independent effect represented by a coding variant.

JT Interval GWAS Identifies 4 Novel Loci

There have been previous reports where SNVs effect QT interval and QRS interval, a measure of ventricular depolarization time, in opposite directions (Arking, Pulit, et al.; Sotoodehnia et al.). While ventricular depolarization time and repolarization time are often co-regulated, the difference in genetic effect indicates this is not universally true. By looking at only ventricular repolarization time, we should have increased statistical power to detect variants that specifically effect ventricular repolarization time and thus detect additional loci while teasing apart the differential regulation of the various phases of ventricular conduction. Ventricular repolarization time begins at the Q point of an EKG; however ventricular depolarization time does not finish until the S point of an EKG. Therefore, the time that both processes are co-occurring is included in the QT interval. We define JT interval

mathematically as subtracting the QRS interval from the QT interval in milliseconds. In ARIC's 15,590 participants, the correlation between QT and JT is 0.92. We analyzed JT interval the same way as QT interval described above, while adding QRS interval as an additional covariate to further remove the effect of ventricular depolarization time on the analysis. Four loci that were not observed as QT loci achieved exome-wide significance, including three loci harboring significant coding variants: *SENP2*, *SLC12A7*, and *NACA*.

in silico Analyses

To further decode the role these new loci might play in regulating ventricular repolarization time, Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT) (Pers et al.) was used to investigate if identified loci contain genes from functional annotated gene sets/pathways. The 45 SNVs from Table 3.1 were used to seed the algorithm, only 38 SNVs were able to be used by the algorithm. Included in Table 3.1 is a list of genes with FDR<5%. Three gene sets passed the false discovery rate (FDR) cutoff of 5%: C1QA subnetwork (ENSG00000173372; p=1.97E-6), fascia adherens (GO:0005916; p=8.28E-6), and ACOT13 subnetwork (ENSG00000112304; p=9.02E-6). Three tissues also passed the FDR cutoff of 5%: Heart Ventricles

(A07.541.560; p=9.56E-4), Heart (A07.541; p=9.74E-4), and Atrial Appendage (A07.541.358.100; p=0.00283).

We also looked up each of the Table 3.1 representative SNVs and GWiS independent SNVs in the GTEx Portal to identify single-tissue expression quantitative trait loci (eQTL). The results are presented in Table 3.1 (left ventricle association noted in bold). Interestingly, rs1361754 was found to be both an exome-wide significant coding variant in *PM20D1* and an eQTL for the same gene in left ventricle. Similarly, rs1805123 in *KCNH2* and rs1042391 in *GMPR*, are both nonsynonymous variants as well as eQTLs, though not in left ventricle.

3.5 Discussion

Our approach of focusing on coding variants has identified 10 novel loci associated with QT/JT interval, and highlighted the role of 17 specific genes, 7 of which are from novel loci: *PM20D1*, *SLC4A3*, *CASR*, *NRAP*, *SENP2*, *SLC12A7*, and *NACA*. Not all 35 loci in the QTIGC study were replicated in our data despite the two studies have comparable discovery population sizes. This is likely due to the ExomeChip focusing on coding variation, with clear lack of coverage for some parts of the genome.

Previous studies have implicated roles for potassium ion regulation, sodium ion regulation, calcium ion regulation, and autonomic

control of QT interval (Porta et al.), and our results provide support for each of these pathways. *SLC12A7* (*KCC4*), which GTEx shows is highly expressed in the left ventricle, is a potassium-chloride cotransporter involved in potassium efflux (Mount et al.). *CASR* is a G protein-coupled receptor that maintains circulating calcium ion homeostasis via PTH secretion in the parathyroid and kidney tubule ion handling (Hendy et al.).

Other genes do not fit into our current understanding of QT interval regulation. *SLC4A3* (*AE3*), which GTEx shows is highly expressed in the heart and brain, has been found to be involved in anion exchange and cytoskeleton structural organization in neurons (Kopito et al.). *SENP2*, as a part of a nuclear pore complex, helps process *SUMO1* (*UBL1*, ubiquitin-like protein) into its conjugatable form (Nishida et al.; Zhang, Saitoh, and Matunis). *SENP2*, which GTEx shows is highly expressed in the testis, brain and to a lesser extent ubiquitously, functional characterization in heart will be needed. Likewise, little is known about *PM20D1* and functional characterization of this gene will be needed.

In addition to previously implicated pathways, our analyses highlight a role for the internal structure of myocytes and interconnection of myocytes in modulating QT interval duration, here after called mechanical control of QT interval. The GO category of fascia adherens (GO:0005916), identified by DEPICT, is comprised of the genes

that code for the structure that links myofibrils between cardiomyocytes and contains N-cadherin in the intercalated disc. The intercalated disc contains fascia adherens, desmosomes, and gap junctions, the last of which is known to play a role in ion-mediated relaying of action potentials between cardiomyocytes and, in combination with the gene *NOS1AP*, has been implicated as regulating QT interval (Kapoor et al.). We further implicate a non-ion dependent structural/mechanical interconnect via the fascia adherens. *NRAP*, found to have a significant independent coding variant, likely anchors terminal actin filaments of myofibrils to other protein complexes beneath the sarcolemma (Luo, Zhang, et al.; Luo, Leroy, et al.). Thus, *NRAP*, which GTEx shows is expressed exclusively in skeletal muscle and heart, likely plays a role in the mechanical control of QT interval. Likewise, skNAC, a muscle-specific isoform of *NACA*, which was found to have a significant independent coding variant, also likely affect mechanical control of QT interval. Most skNAC-specific knockout mice die between embryonic days 10.5 and 12.5 due to cardiac defects, showing interventricular septal defects and a thin myocardial wall (Park et al.). With these three points of evidence combined with the previously known locus and GWiS-implicated gene, *TTN*, a clear class of genes mechanically effecting ventricular repolarization time emerges.

In summary, we have identified 10 loci newly associated with ventricular repolarization time. This brings the total number of ventricular repolarization time associated loci to 45. Additionally, we have directly implicated 17 specific genes contained in these loci as likely affecting ventricular repolarization time and outlined a class of genes that mechanically control QT interval. These new discoveries will likely allow for the development of novel vectors for the prevention of lethal ventricular arrhythmias and SCD.

3.6 Tables

Table 3.1: Investigation of All Ventricular Repolarization Loci

Investigation of 35 previously known ventricular repolarization time loci presented by QTIGC and 10 novel loci from the current efforts. Representative SNVs are chosen to have the lowest p-value in each locus. Variant position information is from build GRCh37. Effect size estimates are in milliseconds (ms). However, p-values were determined from inverse rank normal transformed residuals to avoid p-value inflation from the analysis of rare variants. Loci are considered significantly associated if passing a Bonferroni correction, p-value < 2E-07. CAF = coded allele frequency. N = number of samples. SE = standard error. Discovery column indicates “QTIGC Validated” if a SNV was exome-wide significant

in this study, “QTIGC-only” if the locus could not be replicated, or “Novel” if the locus was first identified in this study. Within the QTIGC Implicated Gene(s) column, evidence for the gene is: c = coding variant, t = eQTL transcript, p = in silico protein-protein interactor, i = immunoprecipitation interactor. DEPICT genes pass FDR<5% cutoff. Expression quantitative trait loci (eQTL) genes are pulled from the GTEx portal using the representative SNV and GWiS independent SNVs. Gene is bold if the eQTL is in the left ventricle.

Nearby Gene	SNV	Chr	Pos	Coded/ Noncoded Allele	CAF	N	Effect in ms (SE)	Pvalue
RNF207	rs709209	1	6,278,414	G/A	0.379	95,626	1.23 (0.09)	1E-48
TCEA3	rs1077514	1	23,766,233	G/A	0.179	92,753	-0.58 (0.11)	4E-08
NOS1AP	rs12143842	1	162,033,890	T/C	0.240	95,626	3.18 (0.10)	3E-255
ATP1B1	rs10919071	1	169,099,483	G/A	0.115	95,626	-1.37 (0.13)	3E-30
SLC8A1	rs2540226	2	39,959,060	T/G	0.482	92,753	0.24 (0.08)	2E-03
SP3	rs1047640	2	174,820,750	C/T	0.120	95,626	0.60 (0.12)	3E-06
TTN-CCDC141	rs72648998	2	179,575,511	T/C	0.054	95,626	1.00 (0.18)	3E-09
SPATS2L	rs192861441	2	201,303,848	A/G	0.004	95,626	-2.22 (0.67)	3E-04
SCN5A-SCN10A	rs12053903	3	38,593,393	C/T	0.379	95,626	-0.88 (0.09)	1E-26
C3ORF75	rs2276853	3	47,282,303	G/A	0.411	95,626	-0.36 (0.08)	2E-05
SLC4A4	rs7689609	4	72,083,374	C/T	0.212	85,380	0.64 (0.12)	4E-08
SMARCAD1	rs7439869	4	95,173,779	T/C	0.378	95,626	0.41 (0.08)	8E-07
GFRA3	rs4835768	5	137,441,767	G/A	0.485	95,626	0.34 (0.08)	7E-05
GMPR	rs1042391	6	16,290,761	T/A	0.551	89,579	-0.42 (0.09)	3E-06
SLC35F1-PLN	rs11153730	6	118,667,522	C/T	0.467	95,626	1.41 (0.08)	5E-74
CAV1	rs3807989	7	116,186,241	A/G	0.429	95,626	0.54 (0.08)	4E-12
KCNH2	rs1805123	7	150,645,534	G/T	0.214	95,626	-1.47 (0.10)	7E-51
NCOA2	rs2926707	8	71,164,680	G/T	0.348	92,753	0.31 (0.09)	3E-04
LAPTM4B	rs17831160	8	99,045,866	A/G	0.030	95,626	-0.64 (0.24)	3E-03
AZIN1	rs143025416	8	104,432,659	A/G	0.001	95,626	4.90 (1.55)	2E-03
GBF1	rs143226354	10	104,174,986	T/C	0.000	95,626	14.18 (4.66)	4E-03
KCNQ1	rs2074238	11	2,484,803	T/C	0.074	89,284	-3.58 (0.16)	8E-130
FEN1-FADS2	rs1535	11	61,597,972	G/A	0.325	95,626	-0.48 (0.09)	8E-10
ATP2A2	rs11068997	12	110,383,141	A/G	0.040	95,626	-0.94 (0.21)	4E-07
KLF12	rs1886512	13	74,520,186	A/T	0.381	80,552	0.57 (0.09)	2E-10
ANKRD9	rs11704	14	102,808,655	C/G	0.291	89,579	0.35 (0.09)	7E-05
USP50-TRPM7	rs8042919	15	50,878,630	A/G	0.097	95,626	-0.57 (0.14)	4E-05
CREBBP	rs143903106	16	3,336,067	T/G	0.001	95,626	4.10 (1.46)	5E-03
LITAF	rs8049607	16	11,691,753	T/C	0.503	95,626	1.05 (0.08)	8E-44
MKL2	rs30208	16	14,428,853	T/C	0.501	95,626	0.45 (0.08)	2E-09
CNOT1	rs7188697	16	58,622,178	G/A	0.247	91,615	-1.57 (0.10)	4E-63
LIG3	rs2074518	17	33,324,382	A/G	0.428	92,753	-0.79 (0.08)	2E-21
PRKCA	rs9912468	17	64,318,357	G/C	0.417	89,579	-0.68 (0.08)	2E-15
KCNJ2	rs17779747	17	68,494,992	T/G	0.304	93,948	-1.08 (0.09)	3E-37
KCNE1	rs727957	21	35,880,072	T/G	0.168	95,626	0.48 (0.11)	3E-05

Function	Discovery	QTIGC Implicated Gene(s)	Gene(s) with independent coding variation	DEPICT Implicated Gene(s)	eQTL
Nonsynonymous	QTIGC Validated	*RNF207(c)*	*RNF207*	*RNF207*	*GPR153*
Intronic	QTIGC Validated	*TCEA3(t)*			***TCEA3,*** *ASAP3*
Intergenic	QTIGC Validated				
Intronic	QTIGC Validated	*ATP1B1(ti), NME7(t)*			*NME7*
Intergenic	QTIGC-only	*SLC8A1(p)*			*THUMPD2*
Nonsynonymous	QTIGC-only		*SP3*		
Nonsynonymous	QTIGC Validated	*CCDC141(i), TTN(i)*	*TTN*		*FKBP7, PRKRA*
Nonsynonymous	QTIGC-only	*SPATS2L(t), SGOL2(p)*			
Intronic	QTIGC Validated	*SCN5A(p)*	*SCN10A*	*SCN5A*	*SNORA6, SCN5A*
Nonsynonymous	QTIGC-only	*KLHL18(t), PTPN23(t), SCAP(t), SETD2(t), MYL3(i)*		*NBEAL2*	*NBEAL2, PTPN23, SCAP*
Intronic	QTIGC Validated				
Nonsynonymous	QTIGC-only		*SMARCAD1*		
Intergenic	QTIGC-only	*FAM13B(t), ETF1(p)*		*MYOT, FAM13B*	
Nonsynonymous	QTIGC-only	*GMPR(c), ATXN1(tp),*	*GMPR*		*GMPR*
Intergenic	QTIGC Validated	*PLN(i)*		*PLN*	*SSXP10*
Intronic	QTIGC Validated	*CAV1(pi), CAV2(pi)*			*AC002066.1*
Nonsynonymous	QTIGC Validated	*KCNH2(p)*	*KCNH2*		*KCNH2*
Intronic	QTIGC-only				
Nonsynonymous	QTIGC-only				
Nonsynonymous	QTIGC-only				
Splicing/ Nonsynonymous	QTIGC-only	*ACTR1A(i)*			
Intronic	QTIGC Validated	*C11ORF21(t), PHEMX(t), TSPAN32(t), KCNQ1(p)*	*KCNQ1*	*KCNQ1*	
Intronic	QTIGC Validated	*FADS1(t), FADS2(t), FADS3(t)*			***FAD2,*** *FADS1****,*** *TMEM258*
Nonsynonymous	QTIGC-only	*VPS29(t), GPN3(t), ARPC3(t), C12ORF24(t), ATP2A2(pi)*	*GIT2, TCTN1*	*ATP2A2, PPTC7*	
Intronic	QTIGC Validated	*KLF12(t)*			***KLF12***
UTR3	QTIGC-only	*ANKRD9(t)*		*ANKRD9*	*ZNF839*
Nonsynonymous	QTIGC-only				*SPPL2A, AP4E1, USP50*
Nonsynonymous	QTIGC-only	*TRAP1(i)*			
Intergenic	QTIGC Validated	*LITAF(t)*			***LITAF***
Intergenic	QTIGC Validated				
Intronic	QTIGC Validated	*NDRG4(t), CNOT1(t), GOT2(i)*			*SETD6, NDRG4*
Intronic	QTIGC Validated	*LIG3(t), CCT6B(t), UNC45B(i)*			***LIG3,*** *CCT6B****, RFFL,*** *RP5-837J1.2*
Intronic	QTIGC Validated	*PRKCA(t)*			***PRKCA***
Intergenic	QTIGC Validated				
Intronic	QTIGC-only	*KCNE1(cp)*			

Nearby Gene	SNV	Chr	Pos	Coded/ Noncoded Allele	CAF	N	Effect in ms (SE)	Pvalue
PM20D1	rs1361754	1	205,801,872	G/A	0.511	95,626	0.47 (0.08)	1E-09
SLC4A3	rs55910611	2	220,500,412	A/G	0.006	74,508	-3.06 (0.61)	2E-07
CASR	rs1801725	3	122,003,757	T/G	0.126	95,626	-0.58 (0.12)	4E-08
ZNF37A	rs4934956	10	38,814,815	T/C	0.497	70,792	0.58 (0.10)	2E-10
NRAP	rs3189030	10	115,393,929	A/G	0.299	95,626	-0.48 (0.09)	4E-08
GOSR2	rs17608766	17	45,013,271	C/T	0.123	95,626	0.72 (0.12)	3E-09
SENP2	rs6762208	3	185,331,165	A/C	0.358	92,046	0.44 (0.08)	2E-07
SLC12A7	rs737154	5	1,065,399	C/T	0.500	92,046	-0.40 (0.08)	2E-07
CDKN1A	rs9470361	6	36,623,379	A/G	0.249	92,046	-0.76 (0.09)	2E-15
NACA	rs2926743	12	57,114,100	A/G	0.252	92,046	0.53 (0.09)	6E-08

Function	Discovery	QTIGC Implicated Gene(s)	Gene(s) with independent coding variation	DEPICT Implicated Gene(s)	eQTL
Nonsynonymous	QT Novel	N/A	*PM20D1*		***PM20D1,*** *NUCKS1,* ***RAB7L1,*** *SLC41A1*
Nonsynonymous	QT Novel	N/A	*SLC4A3*		
Nonsynonymous	QT Novel	N/A	*CASR*		*CSTA*
Intergenic	QT Novel	N/A			
Nonsynonymous	QT Novel	N/A	*NRAP*	*NRAP*	***CASP7***
UTR3	QT Novel	N/A			*RPRML*
Nonsynonymous	JT Novel	N/A	*SENP2*		
Splicing/ Synonymous	JT Novel	N/A	*SLC12A7*		*NKD2*
Intergenic	JT Novel	N/A			
Nonsynonymous	JT Novel	N/A	*NACA*	*RBMS2*	

Table 3.2: Exploration of the *SCN5A-SCN10A* Locus

Table 3.2a lists the 3 exome-wide significant coding variants in the *SCN5A-SCN10A* locus from the all ancestries QT association. Table 3.2b contains the result of running GWiS on all 413 variants in the locus from the European ancestry-only QT association. 4 variants representing 4 independent effects in the locus are shown with one of them being represented by a coding variant in *SCN10A*. "R2" is the r-squared between the SNV being added to the model and the previous SNV held in the model (or zero for the first SNV). "# Tests" is the effective number of independent tests in the locus, which is fewer than "SNPs Tested" due to LD between SNVs.

Gene	SNV	Chr	Pos	Coded/ Noncoded Allele	CAF	N	Effect in ms	Pvalue	Function
DLEC1	rs116202356	3	38,103,776	G/A	0.02	95,626	2.22	3E-11	Nonsynonymous
SCN5A	rs1805124	3	38,645,420	T/C	0.24	95,626	0.66	7E-12	Nonsynonymous
SCN10A	rs6795970	3	38,766,675	A/G	0.37	95,626	-0.67	3E-17	Nonsynonymous

Nearby SNPs

Gene	SNV	Chr	Pos	Tested	# Tests	N	R2	Pvalue	Function
SCN5A	rs12053903	3	38,593,393	413	365.28	83,884	0.000	9E-28	Intronic
SCN5A	rs3922844	3	38,624,253	413	365.28	81,011	0.002	3E-18	Intronic
SCN10A	rs6795970	3	38,766,675	413	365.28	83,884	0.001	2E-16	Nonsynonymous
	rs9851724	3	38,719,935	413	365.28	70,434	0.010	9E-12	Intergenic

Table 3.3: Cohort Demographics

Short Name	AGES	ARIC – EA	ARIC – AA	BRIGHT
Long Name	Age, Gene/Environment Susceptibility (AGES Reykjavik) Study	The Atherosclerosis Risk in Communities Study	The Atherosclerosis Risk in Communities Study	British Genetics of Hypertension
N, participants after exclusion	2381	10246	3567	821
Sex, women, %	61.74	53.85	62.88	60.9
Age, years, mean±SD	76.12±5.405	54.2±5.683	53.37±5.788	57.54±10.65
Age, min-max	66-95	44-66	44-66	22-85
Height, cm, mean±SD	166.4±9.152	168.5±9.394	167.9±8.876	165.9±9.05
BMI, kg/m2, mean±SD	27.09±4.464	26.99±4.862	29.59±6.17	27.42±3.84
Heart rate, bpm, mean±SD	66.45±11.38	66.15±9.815	66.58±11.01	63.97±11.5
QT interval, ms, mean±SD	405.3±34.47	398.8±28.97	400±33.01	421.9±24.4
QT interval, ms, min-max	292-584	288-646	308-696	363-531
JT interval, ms, mean±SD	314±33.49	307.6±28.67	310±32.48	328.4±23.97
JT interval, ms, min-max	214-492	208-556	212-612	266-448
Study design		Population-based	Population-based	Hypertensive cases
Ethnicity and origin		Americans with European Ancestry	Americans with African Ancestry	White Europeans from UK
Exome Chip version	“1.0”	“1.0”	“1.0”	“1.0”
Genotype calling software	centrally at CHARGE	centrally at CHARGE	centrally at CHARGE	GenCall and zCall
Short description on QC	centrally at CHARGE	centrally at CHARGE	centrally at CHARGE	followed Oxford's "ExomeChip_QC_SOP_v5 " protocol
Related individuals (yes/no)?	No	No	No	No
Familial adjustment method (if applicable)	N/A	N/A	N/A	N/A
Population stratification assessment and adjustment	2 PCs	10 PCs	10 PCs	10 PCs
Analysis software version (e.g. seqMeta v1.6.0)	seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.0	R (3.01), seqMeta (1.3)

CAMP	CHS – EA	CHS – AA	ERF	FHS
MGH Cardiology and Metabolic Patient Cohort	The Cardiovascular Health Study	The Cardiovascular Health Study	Erasmus Rucphen Family Study	Framingham Heart Study
2873	3363	648	965	7062
41.7	59.41	64.51	55.23	54.87
61.6±11.4	72.42 (5.43)	72.57 (5.64)	48.14 (14.30)	39.33±9.87
31-81	65-100	65-93	16.65 - 85.27	19-72
171.5±10.1	164.6 (9.36)	164.27 (9.08)	167.60 (9.48)	168.93±9.54
28.75±5.85	26.32 (4.48)	28.48 (5.5)	26.82 (4.59)	26.10±4.98
66.82±12.10	64.36 (10.23)	67.51 (11.49)	62.89 (10.41)	68.97±13.64
417.15±23.00	414.99 (32.22)	407.28 (34.96)	398.82 (28.20)	393.19±36.77
336-574	308-544	312-540	304 - 520	260-610
327.97±24.71	326.16 (31.47)	319.56 (34.87)	301.37 (27.34)	328±30
253-482	212-452	216-456	200 - 408	217-511
Population-based	Population-based	Population-based	Population-based family	Population-based
European Ancestry	Americans with European Ancestry	Americans with African Ancestry	European	Americans with European Ancestry
Infinium HumanCoreExome-24 BeadChips	“1.0”	“1.0”	“1.1”	“1.0”
GeneCall + Zcall	centrally at CHARGE	centrally at CHARGE	zCall	centrally at CHARGE
SNP call rate ≥95%, HWE P≥1E-6	centrally at CHARGE	centrally at CHARGE	using CHARGE recommendations	centrally at CHARGE
No	No	No	Yes	Yes
N/A	N/A	N/A	adjusted for kinship matrix in seqMeta	famSKAT
10 PCs	10 PCs	10 PCs	NA	10 PCs
seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.0

GS:SFHS	GOCHA	GRAPHIC	Inter99
Generation Scotland: Scottish Family Health Study	Genetics of Cerebral Hemorrhage with Anticoagulation	The Genetic Regulation of Arterial Pressure of Humans in the Community	Inter99
9027	360	1736	5695
59.3	195 (54.2)	49.06	52.1
51.83, 13.57	73.2±8.3	39.07±14.52	46.2±7.9
18-80	48-100	18-61	29.7-61.3
167.6, 9.54	168.5±10.5	171.08±9.47	172.2±9.1
26.95, 5.17	26.1±4.6	26.03±4.62	26.3±4.7
69.65, 11.36	68.3±13.7	66.0±10.34	67.0±10.9
405.92, 30.99	428.6±30.6	404.05±19.91	403.5±26.8
304-552	373-667	343-469	310-538
316.49, 30.56	N/A	301.80±27.88	312.5±26.7
216-464	N/A	228-406	228-436
Population-based with families	Population-based	Population-based	Population-based
European-ancestry from Scotland	Americans with European Ancestry	European Caucasian	European
HumanOmniExpressExome8v1-2_A and HumanOmniExpressExome-8v1_A	“1.0”	“1.0”	“1.0”
Beadstudio-Gencall v3.0	Zcall at Broad	gencall/zcall	GenCall and zCall
ID call rate >97%, SNP call rate filter 98%, HWE cutoff <1E-6	Samples excluded missing>2%, mismatch gender, and first- or second-degree relatives identified based on identity-by-descent allele sharing (pi-hat>0.185). Variants call rate<95%, mean heterozygosity >±3sd, departure from Hardy-Weinberg equilibrium at p<1×10-6 in control subjects, or differential missingness in cases and controls were excluded. All SNPs were aligned on the forward strand and coded to the same minor alleles in both datasets.	As in the FINAL.CHARGE-EX.EKG.QTAnalysisPlan	Exclusion criteria: SNPs 1) Call rate >98%; 2) HWE P >10-4; Individuals: 1) Heterozygosity was calculated separately for maf < 1% and maf > 1% and samples were dropped judged by plots; 2) Cryptic relatedness (related to 20 or more individuals).
Yes	No	Yes	No
Kinship matrix	N/A	Kinship	N/A
N/A	2 PCs		10 PCs
seqMeta v1.6.5	seqMeta v1.5.0	seqMeta	seqMeta v1.5

JHS	KORA	CROATIA-Korcula	Lifelines	MESA – EA
The Jackson Heart Study	Kooperative Gesundheitsforschung in der Region Augsburg	CROATIA-Korcula	The Lifelines Cohort Study	Multi-Ethnic Study of Atherosclerosis (MESA) Cohort
2216	2672	295	1943	2324
62.54	52.3	62.4	59.59	53.7
53.06±12.69	48.8±13.1	54.23, 13.38	45.27±13.09	62.36±10.17
21-91	25-74	18-88	18-87	44-84
169.35±9.34	168.3±9.3	168.3, 8.91	174.66±9.32	168.4±9.266
31.34±6.42	27.0±4.4	27.99, 4.24	25.89±4.55	27.66±5.354
68.39±10.14	65.1±10.2	65.85, 9.44	68.65±11.06	66.26±10.14
414.14±31.62	407.5±26.7	401.64, 29.44	393.53±26.89	399.1±30.07
290-580	316-542	270.0-510.0	289-525	334-538
320.37±30.46	315.7±26.6	305.79, 29.36	299.92±26.79	319.32±28.65
212-466	234-442	176.0-408.0	202-415	240-438
Mixed family and population-based	Population-based	Isolate population	Population-based	Population-based
African American	European / Germany	European-ancestry from Croatian island of Korcula	European, Netherlands	
“1.0”	“1.0”	HumanExome-12v1_A	“1.1”	“1.0”
centrally at CHARGE	CHARGE cluster file	Beadstudio-Gencall v3.0	GeneCall + Zcall	centrally at CHARGE
centrally at CHARGE	sample selection based on genCall+zCall following SOP v5	ID call rate >97%, SNP call rate filter 98%, HWE cutoff <1E-6	GeneCall + Zcall; SNP Callrate≥95%, HWE>10-6, Sample Callrate<95%, on PCA outliers, duplicates, gender discordance and mean IBS	centrally at CHARGE
Yes	No	Yes	No	No
Kinship adjustment	N/A	Kinship matrix	N/A	N/A
10 PCs	pairwise exclusion of samples with PI_HAT>0.1875, keep sample with higher callrate	N/A	5 PCs	2 PCs
seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.5	seqMeta v1.6.0	seqMeta v1.6.0

MESA – AA	MESA – HA	MESA – CH	NEO	RS
Multi-Ethnic Study of Atherosclerosis (MESA) Cohort	Multi-Ethnic Study of Atherosclerosis (MESA) Cohort	Multi-Ethnic Study of Atherosclerosis (MESA) Cohort	The Netherlands Epidemiology of Obesity (NEO) Study	The Rotterdam Elderly Study
1501	1382	750	6047	2419
54.9	52.6	51.6	52.03	55.25
62.06±10.03	61.16±10.24	62.21±10.37	56.0±5.94	68.6±8.363
45-84	44-84	44-84	44-66	55-101
168.36±9.52	161.68±9.34	161.49±8.58	173.62±9.59	167.5±9.384
30.1±5.88	29.5±5.15	23.99±3.29	30.05±4.82	26.21±3.591
63.01±10.28	63.5±9.39	63.06±8.63	65.7±11.39	70.73±12.28
410.33±31.73	408.8±29.75	410.97±29.38	406.6±30.5	396.7±29.197
320-512	328-530	334-554	244-666	282-524
319.4±30.99	317.79±29.42	321.67±29.55	312.9±28.7	299.82±28.158
240-420	238-428	256-450	188-484	196-416
Population-based	Population-based	Population-based	Population-based	Population-based
			European Ancenstry from the Netherlands	Europeans with European Ancestry
“1.0”	“1.0”	“1.0”	HumanCoreExome-24v1-0	“1.0”
centrally at CHARGE	centrally at CHARGE	centrally at CHARGE	GenCall	centrally at CHARGE
centrally at CHARGE	centrally at CHARGE	centrally at CHARGE	Outlying individuals were excluded on the basis of relatedness, non-European ancestry, sex discrepancy and heterzygosity	centrally at CHARGE
No	No	No	No	No
N/A	N/A	N/A	N/A	N/A
2 PCs	2 PCs	2 PCs	10 PCs	5 PCs
seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.5	seqMeta v1.6.0

SHIP	TwinsUK	UHP
Study of Health In Pomerania	TwinsUK	Utrecht Health Project
6224	466	1731
52.52	93.56	55
49.56±15.32	52.08±11.65	39.10±12.956
20-82	18-83	18-91
169.54±9.31	163.2±6.930	174.78±9.779
27.57±4.98	26.72±5.328	24.90±3.875
NA	66.63±10.45	64.60±10.612
406.82±28.43	403.2±27.81	403.48±27.422
308-540	308-500	308-512
312.15±29.36	315.2±27.37	306.65±27.054
212-436	228-402	216-402
Population-based	Twin study	Population-based
EA from Germany	European ancestry, individuals from the United Kingdom	Dutch citizens of European Ancestry
“1.0”	“1.0”	1.1
Gencall (Illumina Genome Studio), followed by zCall	Gencall	GenomeStudio and zCall
Samples: Genotype call rate <98%; High heterozygosity and/or implausible high crypted relatedness; IBS clustering, Unexpected duplicates; Sex mismatches Variants: Call rate <95%; HWE p-value < 10E-4	Excluded sampels with callrate < 97%, removed autosomal heterozygosity outliers (+/- 4SD (calculated for variants with MAF <1% and MAF >=1% separately) , gender mismatches, duplicates as established by identity by descent (IBD) analysis, ethnic outliers as determined by combining with 1000 Genomes Project data (PCA), GWAS concordance (when available). Removed variants with call rate < 95% and pHWE< 1x10-6.	Plink v1.07 was used for QC. Samples with missing SNP rate >5% or discordant sex were excluded. Using SNPs with missingness<1%, MAF>5%, Hardy-Weinberg P<0.001, LD-pruned r2>0.2, we removed samples with heterozygosity > 4 standard deviations from the mean, related samples randomly, and samples from non-European descent based on manual inspection of PCA results that were calculated with Eigensoft. SNPs with missing rates>5% or Hardy-Weinberg equilibrium P<0.001 were removed.
No	No	No
N/A	N/A	N/A
10 PCs	10 PCs	1 PC
seqMeta v1.4.0 (seqMeta v1.3.0 for QT analysis)	seqMeta v1.3	seqMeta v1.6.0

WHI – EA	WHI – AA	YFS
The Women's Health Iniative	The Women's Health Iniative	The Cardiovascular Risk in Young Finns Study
13450	1678	1784
100	100	55.72
66.1±6.547	64.55±6.46	41.92±4.98
50-81	50-79	34-49
161.5±6.6.307	161.9±6.708	172.04±9.22
28.70±5.637	31.13±5.85	26.45±4.94
66.56±10.125	66.77±10.84	60.04±9.47
401.399±29.9322	402.23±31.77	415.11±33.54
290-624	310-520	284-636
315.2±29.49	317.13±30.85	324.66±33.47
204-534	218-426	206-564
Population-based	Population-based	Population-based
Americans with European Ancestry	Americans with African Ancestry	Finnish with European Ancestry
“1.0”	“1.0”	CoreExome v1.0
		GenomeStudio
		SNP and sample call-rate 95%, excess heterozygosity, cryptic relatedness, MDS outliers
No	No	No
N/A	N/A	N/A
2 PCs	2 PCs	4 PCs
seqMeta v1.6.0	seqMeta v1.6.0	seqMeta v1.3.0

Table 3.4: Conditional Analyses in European Ancestry-only ARIC

Conditional analyses with this studies representative SNV and QTIGC's representative SNV in the same QT interval model. EC=ExomeChip (this study). Con=conditional analysis. The "Con Survive" column indicates if the ExomeChip SNV or QTIGC SNV or both have effect size estimates the remain about the same in the conditional model. LD calculations are performed in the merged ExomeChip and HapMap-imputed most likely genotype ARIC Europeans dataset with 9,537 samples. Conditional analyses were run in the same ARIC Europeans dataset, however limited to 9,005 individuals due to phenotype exclusions. Effect sizes are in milliseconds. EC Pos column is position from GRCh37.

Nearby Gene	Chr	EC Pos	EC SNV	EC Function	QTIGC SNV	LD b/w SNVs	ARIC EC Effect	ARIC QTIGC Effect	Con ARIC EC Effect	Con ARIC QTIGC Effect	Con Survive
RNF207	1	6,278,414	rs709209	Nonsynonymous	rs846111	0.696	1.11	1.38	0.16	1.23	QTIGC
SP3	2	174,820,750	rs1047640	Nonsynonymous	rs938291	0.218	1.11	0.77	0.75	0.52	Both
TTN-CCDC141	2	179,575,511	rs72648998	Nonsynonymous	rs7561149	0.044	1.29	-0.12	1.30	0.01	EC
SPATS2L	2	201,303,848	rs192861441	Nonsynonymous	rs295140	0.002	-2.98	0.59	-2.78	0.57	Both
C3ORF75	3	47,282,303	rs2276853	Nonsynonymous	rs17784882	0.900	-0.34	-0.41	0.53	-0.91	QTIGC
SMARCAD1	4	95,173,779	rs7439869	Nonsynonymous	rs3857067	0.539	0.63	-0.65	0.33	-0.41	QTIGC
GMPR	6	16,290,761	rs1042391	Nonsynonymous	rs7765828	0.989	-0.50	0.52			
KCNH2	7	150,645,534	rs1805123	Nonsynonymous	rs2072413	0.823	-1.32	-1.43	0.17	-1.58	QTIGC
LAPTM4B	8	99,045,866	rs17831160	Nonsynonymous	rs11779860	0.004	-0.24	-0.33	-0.18	-0.33	Both
AZIN1	8	104,432,659	rs143025416	Nonsynonymous	rs1961102	0.000	15.26	0.53	15.08	0.53	Both
GBF1	10	104,174,986	rs143226354	Splicing/ Nonsynonymous	rs2485376	0.000		-0.28			
ATP2A2	12	110,383,141	rs11068997	Nonsynonymous	rs3026445	0.018	-0.07	0.46	0.11	0.46	QTIGC
USP50-TRPM7	15	50,878,630	rs8042919	Nonsynonymous	rs3105593	0.099	-0.41	0.67	-0.08	0.66	QTIGC
CREBBP	16	3,336,067	rs143903106	Nonsynonymous	rs1296720	0.000	0.63	0.98	0.32	0.97	QTIGC

Table 3.5: All Exome-wide Significant Coding Variants

Table layout similar to Table 3.1. All coding variants that pass a Bonferroni correction of p-value < 2E-07. Part A contains results from the QT interval association and part B contains results from the JT interval association. Pos column is position from GRCh37. Gene column is the gene the variant is in (not a nearby gene or locus label like other tables in this study).

QT: Gene	SNV	Chr	Pos	Coded/ Noncoded Allele	CAF	N	Effect in ms	Pvalue	Function
RNF207	rs709209	1	6,278,414	A/G	0.38	95,626	1.23	1E-48	Nonsynonymous
RNF207	rs846111	1	6,279,370	G/C	0.24	76,129	1.51	2E-46	Nonsynonymous
F5	rs6027	1	169,483,561	T/C	0.05	74,803	-1.27	1E-10	Nonsynonymous
F5	rs6018	1	169,511,878	T/G	0.05	73,622	-1.25	1E-09	Nonsynonymous
F5	rs6033	1	169,521,853	A/G	0.07	95,626	-0.87	1E-08	Nonsynonymous
PM20D1	rs1361754	1	205,801,872	A/G	0.49	95,626	0.47	1E-09	Nonsynonymous
TTN	rs72648998	2	179,575,511	C/T	0.05	95,626	1.00	3E-09	Nonsynonymous
TTN	rs10497520	2	179,644,855	T/C	0.18	92,753	0.63	9E-09	Nonsynonymous
SLC4A3	rs55910611	2	220,500,412	G/A	0.01	74,508	-3.06	2E-07	Nonsynonymous
DLEC1	rs116202356	3	38,103,776	G/A	0.02	95,626	2.22	3E-11	Nonsynonymous
SCN5A	rs1805124	3	38,645,420	T/C	0.24	95,626	0.66	7E-12	Nonsynonymous
SCN10A	rs6795970	3	38,766,675	A/G	0.37	95,626	-0.67	3E-17	Nonsynonymous
CASR	rs1801725	3	122,003,757	G/T	0.13	95,626	-0.58	4E-08	Nonsynonymous
CEP85L	rs3734382	6	118,886,961	G/T	0.26	95,626	-0.63	4E-13	Nonsynonymous
CEP85L	rs3734381	6	118,887,303	T/C	0.46	91,615	-1.00	4E-36	Nonsynonymous
KCNH2	rs1805123	7	150,645,534	T/G	0.21	95,626	-1.47	7E-51	Nonsynonymous
NRAP	rs3189030	10	115,393,929	G/A	0.30	95,626	-0.48	4E-08	Nonsynonymous
NRAP	rs2185913	10	115,410,234	T/C	0.27	95,626	-0.48	2E-07	Nonsynonymous
KCNQ1	rs17215500	11	2,790,111	C/T	0.00	95,626	46.38	1E-11	Stop
CCT6B	rs2230553	17	33,269,648	C/G	0.34	89,579	-0.56	2E-10	Nonsynonymous
CCT6B	rs9635769	17	33,288,363	C/T	0.45	95,626	0.47	3E-08	Nonsynonymous

JT: Gene	SNV	Chr	Pos	Coded/ Noncoded Allele	CAF	N	Effect in ms	Pvalue	Function
RNF207	rs709209	1	6,278,414	A/G	0.38	92,046	0.22	2E-52	Nonsynonymous
RNF207	rs200882245	1	6,279,316	C/T	0.00	92,046	-2.76	2E-08	Nonsynonymous
RNF207	rs846111	1	6,279,370	G/C	0.24	72,859	0.52	6E-50	Nonsynonymous
F5	rs6027	1	169,483,561	T/C	0.05	71,223	-0.57	4E-11	Nonsynonymous
F5	rs6018	1	169,511,878	T/G	0.05	70,404	-0.56	1E-10	Nonsynonymous
F5	rs6037	1	169,513,583	G/T	0.07	72,538	-0.45	8E-08	Synonymous
F5	rs6033	1	169,521,853	A/G	0.07	92,046	-0.51	2E-09	Nonsynonymous
TTN	rs72648998	2	179,575,511	C/T	0.05	92,046	0.36	2E-09	Nonsynonymous
TTN	rs34819099	2	179,628,918	C/T	0.01	92,046	0.87	6E-08	Nonsynonymous
TTN	rs10497520	2	179,644,855	T/C	0.18	89,173	0.10	1E-08	Nonsynonymous
SLC4A3	rs55910611	2	220,500,412	G/A	0.01	70,928	-3.24	6E-08	Nonsynonymous
DLEC1	rs116202356	3	38,103,776	G/A	0.02	92,046	1.33	2E-18	Nonsynonymous
SLC22A14	rs2070492	3	38,357,817	C/T	0.10	92,046	-0.14	4E-08	Nonsynonymous
SCN5A	rs1805124	3	38,645,420	T/C	0.24	92,046	0.12	5E-25	Nonsynonymous

SCN10A	rs12632942	3	38,764,998	A/G	0.25	92,046	0.08	1E-07	Nonsynonymous
SCN10A	rs6795970	3	38,766,675	A/G	0.36	92,046	-0.33	1E-34	Nonsynonymous
SCN10A	rs57326399	3	38,768,300	T/C	0.24	92,046	0.12	2E-10	Nonsynonymous
CASR	rs1801725	3	122,003,757	G/T	0.12	92,046	-0.22	2E-09	Nonsynonymous
SENP2	rs6762208	3	185,331,165	C/A	0.36	92,046	0.04	2E-07	Nonsynonymous
SLC12A7	rs737154	5	1,065,399	C/T	0.50	92,046	-0.10	2E-07	Splicing/Synonymous
CEP85L	rs3734382	6	118,886,961	G/T	0.26	92,046	-0.13	2E-08	Nonsynonymous
CEP85L	rs3734381	6	118,887,303	T/C	0.46	88,035	-0.14	3E-25	Nonsynonymous
KCNH2	rs1805123	7	150,645,534	T/G	0.21	92,046	-0.62	1E-51	Nonsynonymous
NRAP	rs3189030	10	115,393,929	G/A	0.30	92,046	-0.09	3E-08	Nonsynonymous
NRAP	rs2185913	10	115,410,234	T/C	0.27	92,046	-0.07	6E-08	Nonsynonymous
KCNQ1	rs17215500	11	2,790,111	C/T	0.00	92,046	3.99	6E-12	Stop
NACA	rs2958149	12	57,109,792	A/G	0.25	88,035	0.13	8E-08	Nonsynonymous
NACA	rs2926743	12	57,114,100	A/G	0.25	92,046	0.14	6E-08	Nonsynonymous
CCT6B	rs2230553	17	33,269,648	C/G	0.34	86,309	-0.13	5E-09	Nonsynonymous
CCT6B	rs9635769	17	33,288,363	C/T	0.45	92,046	0.09	2E-08	Nonsynonymous

Table 3.6: GWiS Results

Table contains the result of running GWiS on all variants in each locus from the European ancestry-only QT interval association (or JT interval association for the last four). The SNVs are added into the GWiS model in the order they are listed. “R2” is the r-squared between the SNV being added to the model and the previous SNV held in the model (or zero for the first SNV). “# Tests” is the number of independent tests are by the model, which is less than “SNPs Tested” do to LD between SNVs. For the 35 previously identified loci, LD calculations are shown in Table 3.5 between the QTIGC representative SNV and each of the independent representative SNVs picked by GWiS. LD calculations are performed in the merged ExomeChip and HapMap-imputed ARIC Europeans dataset with 9,537 samples. LD is made bold if >0.5. Row color banding in for each locus. Gene column is the gene the variant is in (not a nearby gene or locus label like other tables in this study). The order of the loci is the same as Table 3.1. Pos column is position from GRCh37.

Gene	SNV	Chr	Pos	SNPs Tested	# Tests	N	R2	SNV Pvalue	Function	LD w/ QTIGC
RNF207	rs709209	1	6,278,414	191	175.55	83,884	0.000	1E-50	Nonsynonymous	**0.696**
RNF207	rs200882245	1	6,279,316	191	175.55	83,884	0.001	1E-07	Nonsynonymous	0.001
ASAP3	rs1077514	1	23,766,233	241	208.24	81,011	0.000	3E-09	Intronic	0.060
	rs12143842	1	162,033,890	214	195.06	83,884	0.000	2E-258	Intergenic	**0.989**
NOS1AP	rs4657178	1	162,210,610	214	195.06	83,884	0.055	2E-104	Intronic	0.053
NOS1AP	rs16857031	1	162,112,910	214	195.06	77,837	0.067	2E-70	Intronic	0.047
ATP1B1	rs10919071	1	169,099,483	182	160.22	83,884	0.000	5E-32	Intronic	**0.964**
SLC8A1		2	39,959,060	26	25.38					
SP3	rs1047640	2	174,820,750	79	77.97	83,884	0.000	2E-06	Nonsynonymous	0.218
TTN	rs72648998	2	179,575,511	651	482.16	83,884	0.000	3E-09	Nonsynonymous	0.044
TTN	rs72646869	2	179,446,381	651	482.16	83,884	0.002	7E-07	Nonsynonymous	0.017
TTN	rs16866378	2	179,393,111	651	482.16	83,884	0.001	8E-07	Nonsynonymous	0.009
SPATS2L		2	201,303,848	140	129.2					
SCN5A	rs12053903	3	38,593,393	413	365.28	83,884	0.000	9E-28	Intronic	**0.970**
SCN5A	rs3922844	3	38,624,253	413	365.28	81,011	0.002	3E-18	Intronic	0.002
SCN10A	rs6795970	3	38,766,675	413	365.28	83,884	0.001	2E-16	Nonsynonymous	0.001
	rs9851724	3	38,719,935	413	365.28	70,434	0.010	9E-12	Intergenic	0.000
C3ORF75		3	47,282,303	327	279.96					
SLC4A4	rs7689609	4	72,083,374	160	148.66	75,316	0.000	6E-08	Intronic	**0.673**
SMARCAD1	rs7439869	4	95,173,779	52	45.25	83,884	0.000	1E-06	Nonsynonymous	**0.539**
GFRA3		5	137,441,767	136	124.32					
GMPR	rs1042391	6	16,290,761	61	57.99	77,837	0.000	1E-06	Nonsynonymous	**0.989**
	rs11153730	6	118,667,522	104	95.86	83,884	0.000	1E-76	Intergenic	**0.988**
	rs12210810	6	118,653,204	104	95.86	77,837	0.055	2E-36	Intergenic	0.050
CAV1	rs3807989	7	116,186,241	93	87.79	83,884	0.000	4E-10	Intronic	0.168
KCNH2	rs1805123	7	150,645,534	316	285.07	83,884	0.000	7E-52	Nonsynonymous	**0.823**
	rs4725982	7	150,637,863	316	285.07	83,884	0.087	2E-48	Intergenic	0.088
NCOA2		8	71,164,680	91	87.98					
LAPTM4B		8	99,045,866	117	112.95					
AZIN1		8	104,432,659	81	76					
GBF1		10	104,174,986	187	167.45					
KCNQ1	rs2074238	11	2,484,803	210	197.04	77,542	0.000	3E-127	Intronic	0.021
KCNQ1	rs12296050	11	2,489,342	210	197.04	83,884	0.018	8E-64	Intronic	**0.992**
KCNQ1	rs17215500	11	2,790,111	210	197.04	83,884	0.000	5E-09	Stop	0.000
KCNQ1	rs800336	11	2,473,131	210	197.04	81,011	0.018	4E-18	Intronic	0.000
FADS2	rs1535	11	61,597,972	439	394.13	83,884	0.000	5E-10	Intronic	**0.952**
GIT2	rs11068997	12	110,383,141	182	166.36	83,884	0.000	6E-07	Nonsynonymous	0.018

TCTN1	rs75714509	12	111,080,097	182	166.36	83,884	0.000	2E-06	Nonsynonymous	0.018
KLF12	rs1886512	13	74,520,186	25	23.82	68,810	0.000	8E-11	Intronic	**0.955**
ANKRD9		14	102,808,655	154	136.31					
USP50-TRPM7		15	50,878,630	177	165.09					
CREBBP		16	3,336,067	507	450.41					
	rs8049607	16	11,691,753	173	159.82	83,884	0.000	1E-41	Intergenic	**0.736**
	rs30208	16	14,428,853	63	60.45	83,884	0.000	2E-11	Intergenic	0.342
CNOT1	rs7188697	16	58,622,178	196	184.33	80,521	0.000	8E-65	Intronic	**0.984**
LIG3	rs2074518	17	33,324,382	281	255	81,011	0.000	4E-20	Intronic	**0.994**
PRKCA	rs9912468	17	64,318,357	95	88.61	77,837	0.000	3E-12	Intronic	**0.993**
	rs17779747	17	68,494,992	27	26.89	83,884	0.000	2E-37	Intergenic	0.388
KCNE1		21	35,880,072	129	118.75					

Gene	SNV	Chr	Pos	SNPs Tested	# Tests	N	R2	SNV Pvalue	Function	LD w/ QTIGC
PM20D1	rs1361754	1	205,801,872	228	206.79	83,884	0.000	4E-10	Nonsynonymous	N/A
SLC4A3	rs55910611	2	220,500,412	500	436.51	64,444	0.000	2E-07	Nonsynonymous	N/A
STK11IP	rs620698	2	220,466,199	500	436.51	81,011	0.000	9E-07	Intronic	N/A
CASR	rs1801725	3	122,003,757	322	275.41	83,884	0.000	5E-08	Nonsynonymous	N/A
	rs4934956	10	38,814,815	30	26.77	61,376	0.000	3E-11	Intergenic	N/A
NRAP	rs3189030	10	115,393,929	232	204.8	83,884	0.000	7E-08	Nonsynonymous	N/A
GOSR2	rs17608766	17	45,013,271	117	108.08	83,884	0.000	5E-09	UTR3	N/A
SENP2	rs6762208	3	185,331,165	156	142.15	80,330	0.000	2E-10	Nonsynonymous	N/A
SLC12A7	rs737154	5	1,065,399	272	245.21	80,330	0.000	1E-06	Splicing/Synonymous	N/A
	rs9470361	6	36,623,379	223	207.83	80,330	0.000	5E-15	Intergenic	N/A
NACA	rs2926743	12	57,114,100	554	482.89	80,330	0.000	2E-08	Nonsynonymous	N/A

3.7 Methods

Genotyping and Quality Control

Followed ExomeChip best practices put out by the CHARGE Consortium (Grove et al.).

Association Analyses and Meta-Analysis

All cohorts excluded individuals with QRS intervals greater than or equal to 120ms, heart rate less than 40 beats per minute (bpm) or greater than 120bpm, left or right bundle branch block, atrial fibrillation on baseline EKG, Wolff–Parkinson–White syndrome (WPW), pacemaker, take class I and class III blocking medication, or are pregnant. Clinical characteristics summary statistics for each cohort are provided in Table 3.3.

Effect size estimates are determined via standard inverse variance weighted meta-analysis on a linear association model with ventricular repolarization time as the outcome using covariates: Age, Sex, RR interval, Height, BMI, and cohort specific adjustments (principal components, clinic, family structure). Significance (p-value) is determined by first inverse rank normal transforming residuals from a linear model with ventricular repolarization time as the outcome using covariates: Age, Sex, RR interval, Height, and BMI, then running a

standard inverse variance weighted meta-analysis on a linear association model with the transformed residuals as the outcome using cohort specific adjustments as covariates. These two models are used in tandem to avoid p-value inflation from the analysis of the rare variants on the ExomeChip while maintaining the easy interpretation of effect sizes in milliseconds.

Representative single nucleotide variants (SNVs) have the lowest p-value in each locus. QT loci are considered discovered if passing a Bonferroni correction, p-value < 0.05 / 209,835 SNVs (2E-07). JT loci are considered discovered if passing a Bonferroni correction, p-value < 0.05 / 208,700 SNVs (2E-07). The difference in the number of SNVs is due to the fact not all cohorts that contributed data to the QT analysis contributed data to the JT analysis. Cohorts contribute slightly different number of SNVs do to individual QC efforts.

LD Calculations and Conditional Analyses

LD calculations are performed in the merged ExomeChip and HapMap-imputed ARIC Europeans dataset with 9,537 samples.

Utilization of Variants to Implicate Genes using GWiS

Gene-Wide Significance (GWiS) (Huang et al.) tests are performed by defining each locus as a “gene” and running on European-only

summery statistics from 22 cohorts for a sample size of 83,884 in QT analyses and 80,330 in JT analyses. GWiS finds the number of independent effects in each locus along with a SNV that best represents each independent effect. This is important because even coding variants are effected by LD causing them to appear significant in the analysis due to the true variant driving the association being nearby. The LD information needed by GWiS was estimated in the ARIC Europeans dataset. An attempt to replace non-coding variants with coding variants in r2>0.8 LD, however this yielded no substitutions.

GTEx eQTL Lookup

We looked up each of the Table 3.1 representative SNVs and GWiS independent SNVs (60 SNVs) in the GTEx Portal for single-tissue expression quantitative trait loci (eQTL). All genes passed FDR<5%. The result is a list of nearby genes that the loci the SNV is in is known to effect the mRNA expression of (Table 3.1). If the expression effect was observed in left ventricle, we made the gene name bold in the table. Genes were excluded if the SNV was towards the bottom of an LD significance peak: *ATP1B1*, *ANKRD9*, *BAZ2A* from the *NACA* locus.

3.8 Cohort Specific Methods

AGES

In anticipation of the sequencing of the human genome and description of the human proteome, the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik) (Tamara B Harris et al., "Age, Gene/Environment Susceptibility-Reykjavik Study") was initiated in 2002. AGES-Reykjavik was designed to examine risk factors, including genetic susceptibility and gene/environment interaction, in relation to disease and disability in old age. The study is multidisciplinary, providing detailed phenotypes related to the cardiovascular, neurocognitive (including sensory), and musculoskeletal systems, and to body composition and metabolic regulation. Relevant quantitative traits, subclinical indicators of disease, and medical diagnoses are identified by using biomarkers, imaging, and other physiologic indicators. The AGES-Reykjavik sample is drawn from an established population-based cohort, the Reykjavik Study. This cohort of men and women born between 1907 and 1935 has been followed in Iceland since 1967 by the Icelandic Heart Association. The AGES-Reykjavik cohort, with cardiovascular risk factor assessments earlier in life and detailed late-life phenotypes of quantitative traits, will create a comprehensive study of aging nested in a relatively genetically homogeneous older population. This approach should facilitate identification of genetic factors that contribute to healthy aging as well as the chronic conditions common in old age.

ARIC

The Atherosclerosis Risk in Communities study (http://www.cscc.unc.edu/aric/) includes 15,792 men and women from four communities in the United States (Jackson, Mississippi; Forsyth County, North Carolina; Washington County, Maryland; suburbs of Minneapolis, Minnesota) enrolled in 1987–1989 and prospectively followed. EKGs were recorded at baseline using MAC PC EKG machines (Marquette Electronics) and processed initially by the Dalhousie EKG program in a central laboratory at the EPICORE Center (University of Alberta). Processing was later repeated for the present study using the GE Marquette 12-SL program (2001 version) at the EPICARE Center (Wake Forest University). All EKGs were visually inspected for technical errors and inadequate quality ("The Atherosclerosis Risk in Communities (ARIC) Study").

BRIGHT

The BRIGHT study (Caulfield et al.) includes 2000 unrelated white European hypertensive individuals. Twelve-lead EKG recordings (Siemens-Sicard 440; http://www.brightstudy.ac.uk/info/sop04.html) producing automated measurements of the JT and QT interval were available for all subjects. All data were subsequently transferred from

each recruitment centre by electronic modem to electrophysiologists from the West of Scotland Primary Prevention Study (Professor Peter MacFarlane) for central reporting.

CAMP

The MGH Cardiology and Metabolic Patient Cohort is comprised of 3850 subjects recruited from the ambulatory MGH Cardiology Practice between 2009 and 2012.

CHS

The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four field centers (Fried et al.). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons were enrolled for a total sample of 5,888. CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

ERF

The Erasmus Rucphen Family study (Pardo et al.) is comprised of a family-based cohort embedded in the Genetic Research in Isolated Populations (GRIP) program in the southwest of the Netherlands. The aim of this program is to identify genetic risk factors for the development of complex disorders. In ERF, twenty-two families that had a large number of children baptized in the community church between 1850 and 1900 were identified with the help of detailed genealogical records. All living descendants of these couples, and their spouses, were invited to take part in the study. Comprehensive interviews, questionnaires, and examinations were completed at a research center in the area; approximately 3,200 individuals participated. Examinations included 12 lead EKG measurements. EKGs were recorded on ACTA electrocardiographs (ESAOTE, Florence, Italy) and digital measurements of the QT and JT intervals were made using the Modular EKG Analysis System (MEANS). Data collection started in June 2002 and was completed in February 2005. In the current analyses, 965 participants for whom complete phenotypic, genotypic and genealogical information was available were studied.

FHS

The objective of the Framingham Heart Study was to identify the common factors or characteristics that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke. The researchers recruited 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, and began the first round of extensive physical examinations and lifestyle interviews that they would later analyze for common patterns related to CVD development. Since 1948, the subjects have continued to return to the study every two years for a detailed medical history, physical examination, and laboratory tests, and in 1971, the Study enrolled a second generation - 5,124 of the original participants' adult children and their spouses - to participate in similar examinations. In 1994, the need to establish a new study reflecting a more diverse community of Framingham was recognized, and the first Omni cohort of the Framingham Heart Study was enrolled. In April 2002 the Study entered a new phase, the enrollment of a third generation of participants, the grandchildren of the Original Cohort. In 2003, a second group of Omni participants was enrolled.

Generation Scotland

The Generation Scotland: Scottish Family Health Study (GS:SFHS) (Smith et al.) is a collaboration between the Scottish Universities and the NHS, funded by the Chief Scientist Office of the Scottish Government. GS:SFHS is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from ~24,000 volunteers, aged 18-98 years, in ~7,000 family groups. Participants were recruited across Scotland, with some family members from further afield, from 2006 - 2011. Most (87%) participants were born in Scotland and 96% in the UK or Ireland. GS:SFHS operates under appropriate ethical approvals, and all participants gave written informed consent.

GOCHA

The Genetics of Cerebral Hemorrhage on Anticoagulation (GOCHA) study is a multicenter study comprised of patients age > 55 years presenting to participating hospitals with primary ICH. Controls were enrolled from ambulatory clinics in the same centers from which cases were recruited (Genes for Cerebral Hemorrhage on Anticoagulation (GOCHA) Collaborative Group).

GRAPHIC

The GRAPHIC Study (Tobin et al.) comprises 2024 individuals from 520 nuclear families recruited from the general population in Leicestershire, UK between 2003-2005 for the purpose of investigating the genetic determinants of blood pressure and related cardiovascular traits. 2 Families were included if both parents aged 40-60 years and two offspring ≥18 years wished to participate. A detailed medical history was obtained from study subjects by standardized questionnaires and clinical examination was performed by research nurses following standard procedures. Measurements obtained included height, weight, waist-hip ratio, clinic and ambulatory blood pressure and a 12-lead EKG.

Inter99

The Inter99 study (Jørgensen et al.) carried out in 1999-2001 included invitation of 12934 persons aged 30-60 years drawn from an age- and sex-stratified random sample of the population. The baseline participation rate was 52.5%, and the study included 6784 persons. The Inter99 study was a population-based randomized controlled trial (CT00289237, ClinicalTrials.gov) and investigated the effects of lifestyle intervention on CVD. Here 5827 participants with information on lipids and exome chip were analyzed. EKG information was obtained from the

MUSE Cardiology Information System (GE Healthcare, Wauwatosa, Wisconsin) analyzed by Marquette 12SL algorithm version 21.

JHS

The Jackson Heart Study (Taylor et al.) (https://www.jacksonheartstudy.org/) includes 5,306 African-American men and women from the three counties, Hinds, Madison, and Rankin, that comprise the Jackson, MS metropolitan area. Participants were enrolled in 2000-2004 and have been followed prospectively. A supine 12-lead digital EKG was recorded with the Marquette MAC/PC digital EKG recorder (Marquette Electronics, Milwaukee, Wis), and with electrode placement that duplicates that of the ARIC study. The EKGs are analyzed in accordance with the Minnesota Code Classification system, via an extensively validated computer algorithm that was developed specifically for epidemiologic studies. In-hospital surveillance EKGs are read visually according to the Minnesota Code Classification system.

KORA

KORA (Kooperative Gesundheitsforschung in der Region Augsburg) (Holle et al.; Wichmann et al.) is a series of independent population-based epidemiological surveys and follow-up studies of participants living in the

city of Augsburg, Southern Germany, or its two adjacent counties. All participants are residents of Germany and have been sampled in strata of age and sex from the local registries. In the baseline survey KORA S4, 4,261 subjects have been examined. 3,080 subjects participated in a 7-year follow-up examination of S4 in 2006-2008. Illumina HumanExome BeadChip was measured in KORA F4 participants.

CROATIA-Korcula

The CROATIA-Korcula (Zemunik et al.) study sampled Croatians from the Adriatic island of Korcula, between the ages of 18 and 88. The fieldwork was performed in 2007 in the eastern part of the island, targeting healthy volunteers from the town of Korčula and the villages of Lumbarda, Žrnovo and Račišće.

Lifelines

LifeLines (Scholtens et al.) is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on

multi-morbidity and complex genetics. Details of the protocol have been described elsewhere (https://www.lifelines.nl/lifelines-research/news). Standard 12-lead EKGs were recorded with CardioPerfect equipment (Cardio Control; currently Welch Allyn, Delft, The Netherlands) and digital measurements of the QT intervals were extracted.

MESA

The Multi-Ethnic Study of Atherosclerosis (MESA) (Bild et al.; Grove et al.) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. The cohort is a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84. Approximately 38 percent of the recruited participants are white, 28 percent African-American, 22 percent Hispanic, and 12 percent Asian (predominantly of Chinese descent). Participants were recruited during 2000-2002 from 6 field centers across the U.S. (at Wake Forest University; Columbia University; Johns Hopkins University; the University of Minnesota; Northwestern University, and the University of California – Los Angeles). All underwent extensive initial physical examination and evaluation. The first examination was followed by 4 examination periods that were 17-20 months long.

NEO

The Netherlands Epidemiology of Obesity (NEO) study (de Mutsert et al.): The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panel. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012.

Participants are currently being followed for the incidence of obesity-related diseases and mortality.

RS

The Rotterdam Elderly Study (Hofman, Brusselle, et al.) is a prospective cohort study in the Ommoord district in the city of Rotterdam, the Netherlands. Following the pilot in 1989, recruitment started in January 1990. The main objectives of the Rotterdam Study were to investigate the risk factors of cardiovascular, neurological, ophthalmological and endocrine diseases in the elderly. Up to 2008, approximately 15,000 subjects aged 45 years or over have been recruited. Participants were interviewed at home and went through an extensive set of examinations, bone mineral densiometry, including sample collections for in-depth molecular and genetic analyses. Examinations were repeated every 3-4 years in potentially changing characteristics. Participants were followed for the most common diseases in the elderly, including coronary heart disease, heart failure and stroke, Parkinson's disease, Alzheimer's disease and other dementias, depression and anxiety disorders, macular degeneration and glaucoma, diabetes mellitus and osteoporosis.

SHIP

The Study of Health In Pomerania (Völzke et al.) is a prospective longitudinal population-based cohort study in Western Pomerania assessing the prevalence and incidence of common diseases and their risk factors. SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations.

TwinsUK

TwinsUK (Moayyeri et al.) is a nation-wide registry of volunteer twins in the United Kingdom, with about 12,000 registered twins (83% female, equal number of monozygotic and dizygotic twins, predominantly middle-aged and older). Over the last 20 years, questionnaire and blood/urine/tissue samples have been collected on over 7,000 subjects,

as well as three comprehensive phenotyping assessments in the clinical facilities of the Department of Twin Research and Genetic Epidemiology, King's College London. The primary focus of study has been the genetic basis of healthy aging process and complex diseases, including cardiovascular, metabolic, musculoskeletal, and ophthalmologic disorders. Alongside the detailed clinical, biochemical, behavioral, and socio-economic characterization of the study population, the major strength of TwinsUK is availability of several 'omics' technologies for the participants. These include genome-wide scans of single nucleotide variants, next-generation sequencing, exome sequencing, epigenetic markers (MeDIP sequencing), gene expression arrays and RNA sequencing, telomere length measures, metabolomic profiles, and gut flora microbiomics.

UHP

The Utrecht Health Project (UHP) (Grobbee et al.) is an ongoing dynamic population study initiated in a newly developed large residential area in Leidsche Rijn, part of the city of Utrecht. All new inhabitants were invited by their general practitioner to participate in the UHP. Written informed consent was obtained and an individual health profile (IHP) was made by dedicated research nurses. The UHP study was approved by the Medical Ethical Committee of the University Medical Center, Utrecht, The

Netherlands. A large number of measures were taken, including anthropomorphic and blood pressure measurements, and each participant filled out a questionnaire. A 12-lead EKG was made at rest and digitally stored. PR, QRS, QT, and RR intervals were calculated automatically.

WHI

The Women's Health Initiative (WHI) ("Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group"; Anderson et al.) is a long-term national health study that has focused on The strategies for preventing heart disease, breast and colorectal cancers, and osteoporotic fractures in postmenopausal women (ref1, ref2). The WHI was designed as a set of randomized controlled clinical trials (CTs) and an observational study (OS). The CT (n = 68,132) included 3 overlapping components: the hormone therapy trials (n = 27,347), dietary modification trial (n = 48,835), and calcium and vitamin D trial (n = 36,282). Eligible women could be part of several of the CT components. Women who were ineligible or unwilling to join the CT were invited to join the OS (n = 93,676). All participants in the CT were administered EKGs every three years. In the current paper we include the baseline EKGs of women who were genotyped on the exomechip.

YFS

The YFS (Raitakari et al.) is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

Chapter 4: Metabochip and PROSe-ICD

4.1 Background

SCD is death caused by loss of heart function in the absence of clinical features that would bring a victim to medical attention. Importantly, SCD lacks symptoms 24 hours before the incident, however SCD does have risk factors that can be evaluated. There are between 180,000 and 450,000 cases of SCD in the United States of America annually (Deo and Albert). ICDs can be used to prevent SCD in patients with systolic heart failure. The World Society of Arrhythmia found that there were 133,262 ICD implants in the United States in 2009 alone (Mond and Proclemer). However, ICDs administer therapy in only a minority of patients, about 20% in our data. The cost assessed for reimbursement is approximately $40,000 per device in 2006 (Stevenson). This represents a significant shortcoming in the clinical selection criteria for patients at greatest risk for SCD. Currently the primary clinical metric used to determine if a patient should receive an ICD is low left ventricular ejection fraction. We wish to find out if we can use genetics to predict if an ICD will be useful for a patient.

4.2 Population

We use data from the PRospective Observational Study of Implantable Cardioverter-Defibrillators (PROSe-ICD) cohort (Cheng et al.). This population consists of patients receiving ICDs for primary

prevention of SCD from one of four centers: the Johns Hopkins Hospital and Bayview Medical Center, Baltimore, MD (JHU), the University of Maryland Hospital, Baltimore, MD (UMD), the Virginia Commonwealth University Hospital, Richmond, VA (VCU), or the Washington Hospital Center, Washington, DC (WHC). Inclusion in the study required a history of acute myocardial infarction, non-ischemic left ventricle dysfunction for at least 9 months, left ventricle ejection fraction less than or equal to 35%, and have an ICD. These patients were followed every 6 months since their operation in person or by phone. Blood samples were collected at each in person follow-up. Patients were evaluated for six primary phenotypes: all-cause mortality (death), all-cause mortality censored at first appropriate ICD shock (death2), appropriate ICD shock (appshock), appropriate ICD therapy (apptherapy), inappropriate ICD shock (inappshock), and inappropriate ICD therapy (inapptherapy). Appropriateness is defined as the first incident adjudicated by two of three cardiologists as lethal without intervention. ICD therapy is defined as antitachycardia pacing or defibrillation shock. Baseline 12-lead Bazett corrected QT interval was measured for all patients. Demographics summarized in Table 4.1.

Table 4.1: PROSe-ICD Demographics

Demographic information for the PROSe-ICD cohort. For all individuals and broken down by ethnicity, EA=European-ancestry, AA=African Americans.

Phenotype	**All**	**EA**	**AA**
Patients N (%)	1066 (100)	638 (59.85)	428 (40.15)
Female N (%)	284 (26.64)	139 (21.79)	145 (33.88)
Age years mean±sd	60.76±12.71	63.63±11.68	56.48±12.98
Age min-max	20.48-85.61	26.06-85.61	20.48-83.33
Center JHU UMD VCU WHC Number	690 80 68 228	479 34 27 98	211 46 41 130
Ischemic Cardiomyopathy N (%)	580 (54.41)	411 (64.42)	169 (39.49)
Non-ischemic Cardiomyopathy N (%)	519 (48.69)	249 (39.03)	270 (63.08)
BMI kg/m2 mean±sd	29.8±6.557	29.28±5.906	30.57±7.36
Height cm mean±sd	173.4±9.458	174.3±9.069	172.1±9.878
Death N (%)	374 (35.08)	238 (37.3)	136 (31.78)
Time to Death years mean±sd	6.386±3.092	6.467±3.149	6.266±3.005
Death2 N (%)	318 (29.83)	199 (31.19)	119 (27.8)
Time to Death2 years mean±sd	5.848±3.206	5.843±3.267	5.856±3.116
Appshock N (%)	157 (14.73)	105 (16.46)	52 (12.15)
Time to Appshock years mean±sd	4.598±3.003	4.718±3.131	4.419±2.796
Apptherapy N (%)	217 (20.36)	133 (20.85)	84 (19.63)
Time to Apptherapy years mean±sd	4.384±3.001	4.558±3.156	4.125±2.736
Inappshock N (%)	123 (11.54)	70 (10.97)	53 (12.38)
Time to Inappshock years mean±sd	4.479±3.06	4.7±3.194	4.149±2.819
Inapptherapy N (%)	155 (14.54)	86 (13.48)	69 (16.12)
Time to Inapptherapy years mean±sd	4.395±3.051	4.636±3.188	4.037±2.801
QTc mean±sd	458.9±43.51	457.3±41.22	461.3±46.68
QTc min-max	263-696	301-661	263-696

4.3 Genotyping and Imputation

All patients were genotyped using the Illumina Cardio-MetaboChip (metabochip) containing over 200,000 SNPs associated with metabolic, atherosclerotic, and/or cardiovascular disease traits (Voight et al.). Associated SNPs were contributed by the following 7 large consortiums: CARDIoGRAM (coronary artery disease), DIAGRAM (type 2 diabetes), GIANT (height and weight), MAGIC (glycemic traits), Lipids (lipids), ICBP-GWAS (blood pressure), and QTIGC (QT interval). Standard QC was performed using Plink (Purcell et al.) removing SNPs with greater than 5% missingness and samples with greater than 10% missingness, miss-matched sex, contaminated samples defined by over representation in nearest neighbor test, and genetic outliers defined by separation from HapMap3 reference population in a PCA plot (miss-matched ethnicity).

This genotyping data was then used to seed an imputation to the 1000G phase3 reference using the ShapeIT (Delaneau et al.) and Impute2 (Howie, Donnelly, and Marchini) software. Unfiltered, this resulted in over 81 million genetic variants typed for each patient. However, using all these variants is not our intention. The metabochip focuses on certain regions of the genome while neglecting others due to its focus on metabolic, atherosclerotic, and/or cardiovascular disease traits. This makes for a low quality imputation in those neglected regions, which in this case is most of the genome. Filtering will be

required. Our intention behind imputing is to allow for easy meta-analyses with other cohorts that were genotyped with other genotyping chips. See section 4.6 for more details.

4.4 Association Model

Since this is a longitudinal cohort with time to event data, the CoxPH models is used to determine if a genetic variants effect the phenotype. Covariates include age, sex, center, whether the instigating cardiomyopathy was ischemic, and 4 Principal Components (PCs) to account for population substructure. The software FAST is used to run this model (Chanda et al.). For the QTc phenotype, a standard linear/additive association model is used, including covariates above and additionally height and BMI. This model was also run using the FAST software.

4.5 Results

Due to the limitations discussed in section 4.3 above, results shown below are filtered for FAST's calculated effective sample size greater than or equal to 100 individuals. Displayed are Manhattan and QQ plots for the 6 primary phenotypes and QTc for European-ancestry and African American individuals (Figures 4.1-4.14).

Figure 4.1: PROSe-ICD AA Appshock FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of appropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

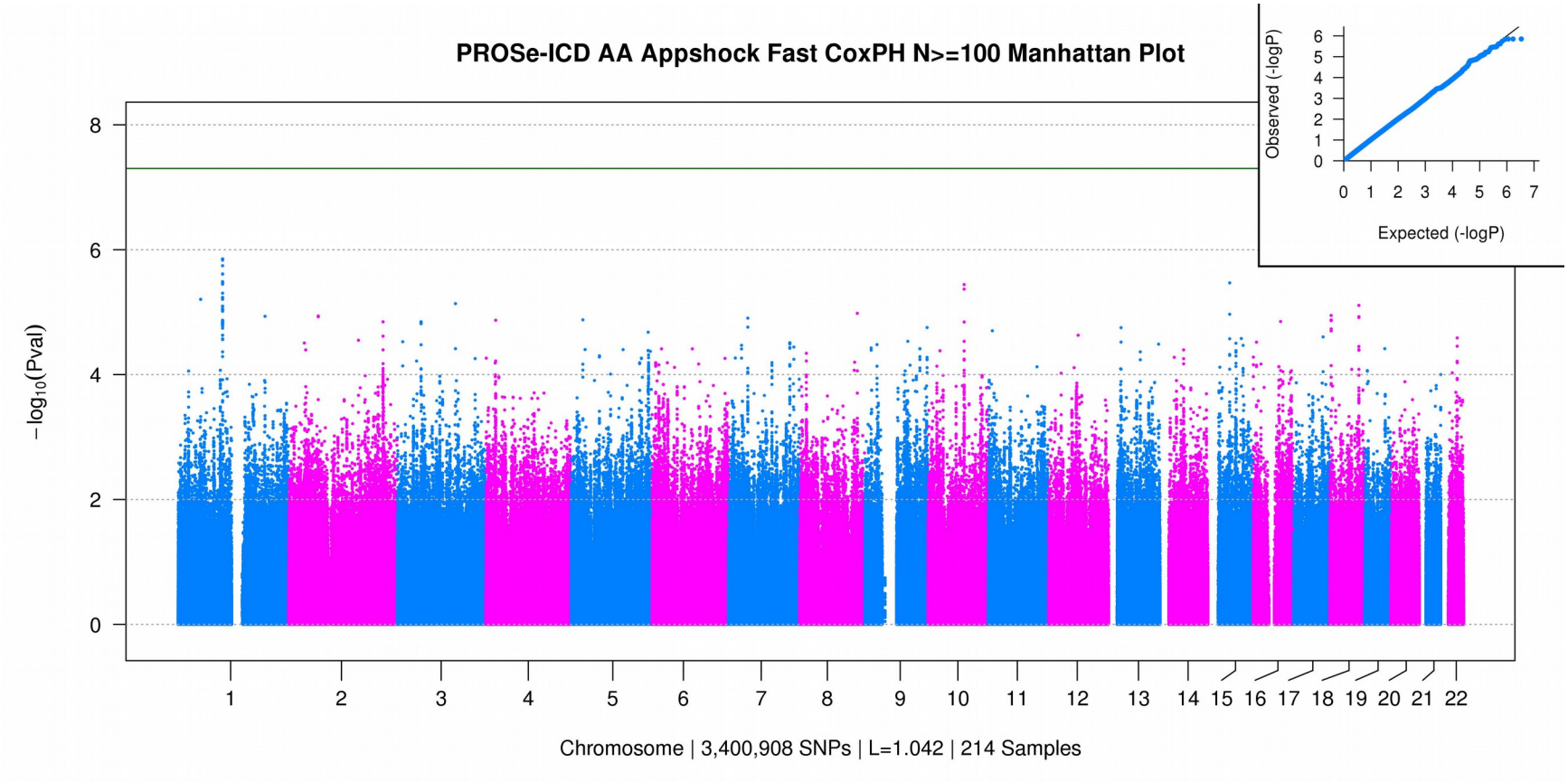

PROSe-ICD AA Appshock Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 3,400,908 SNPs | L=1.042 | 214 Samples

Figure 4.2: PROSe-ICD AA Apptherapy FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of appropriate ICD therapy. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

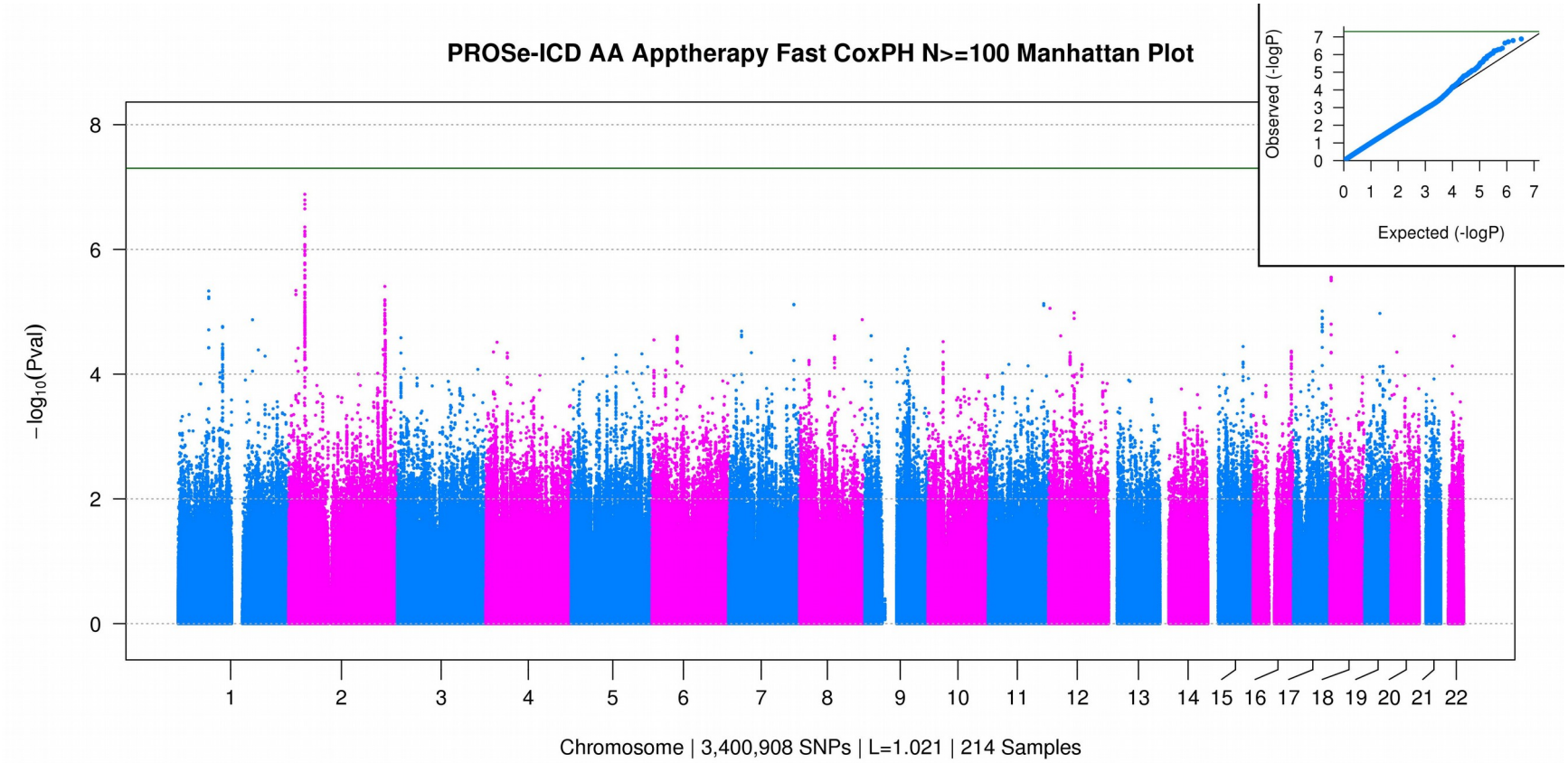

PROSe-ICD AA Apptherapy Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 3,400,908 SNPs | L=1.021 | 214 Samples

Figure 4.3: PROSe-ICD AA Death2 FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of all-cause mortality, censored at first appropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper left. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

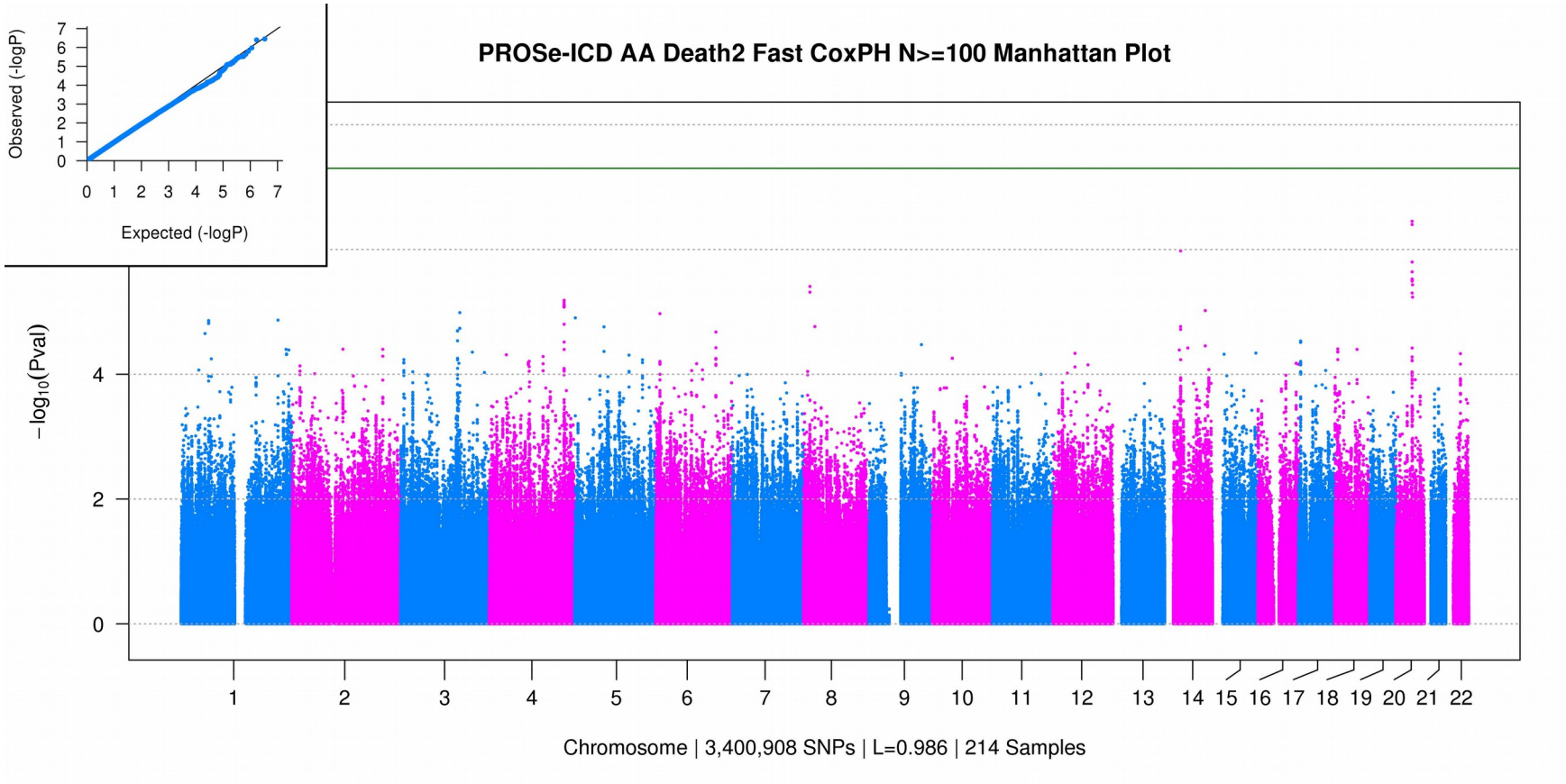

PROSe-ICD AA Death2 Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 3,400,908 SNPs | L=0.986 | 214 Samples

Figure 4.4: PROSe-ICD AA Death FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of all-cause mortality. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

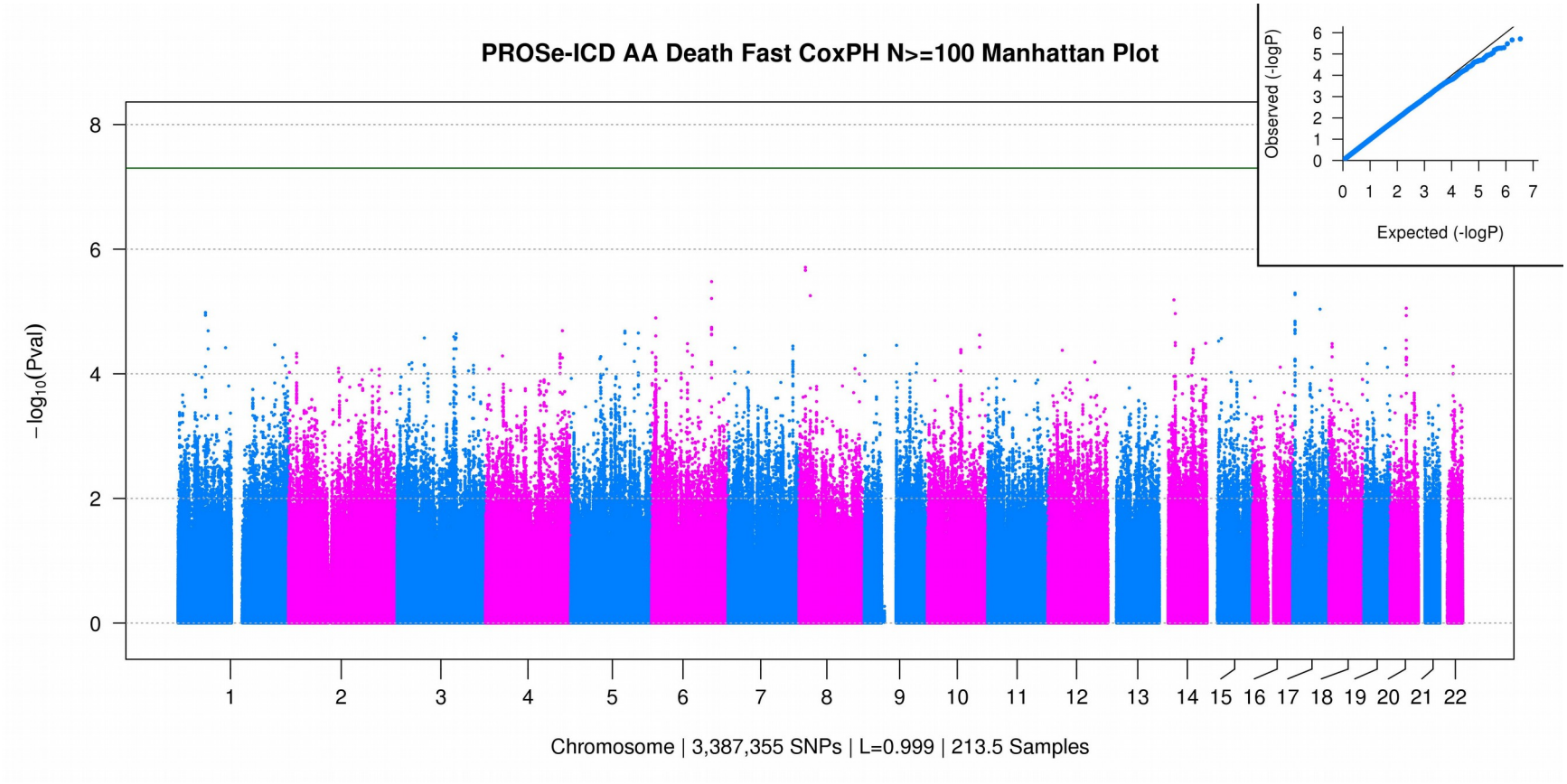

PROSe-ICD AA Death Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 3,387,355 SNPs | L=0.999 | 213.5 Samples

Figure 4.5: PROSe-ICD AA Inappshock FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of inappropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

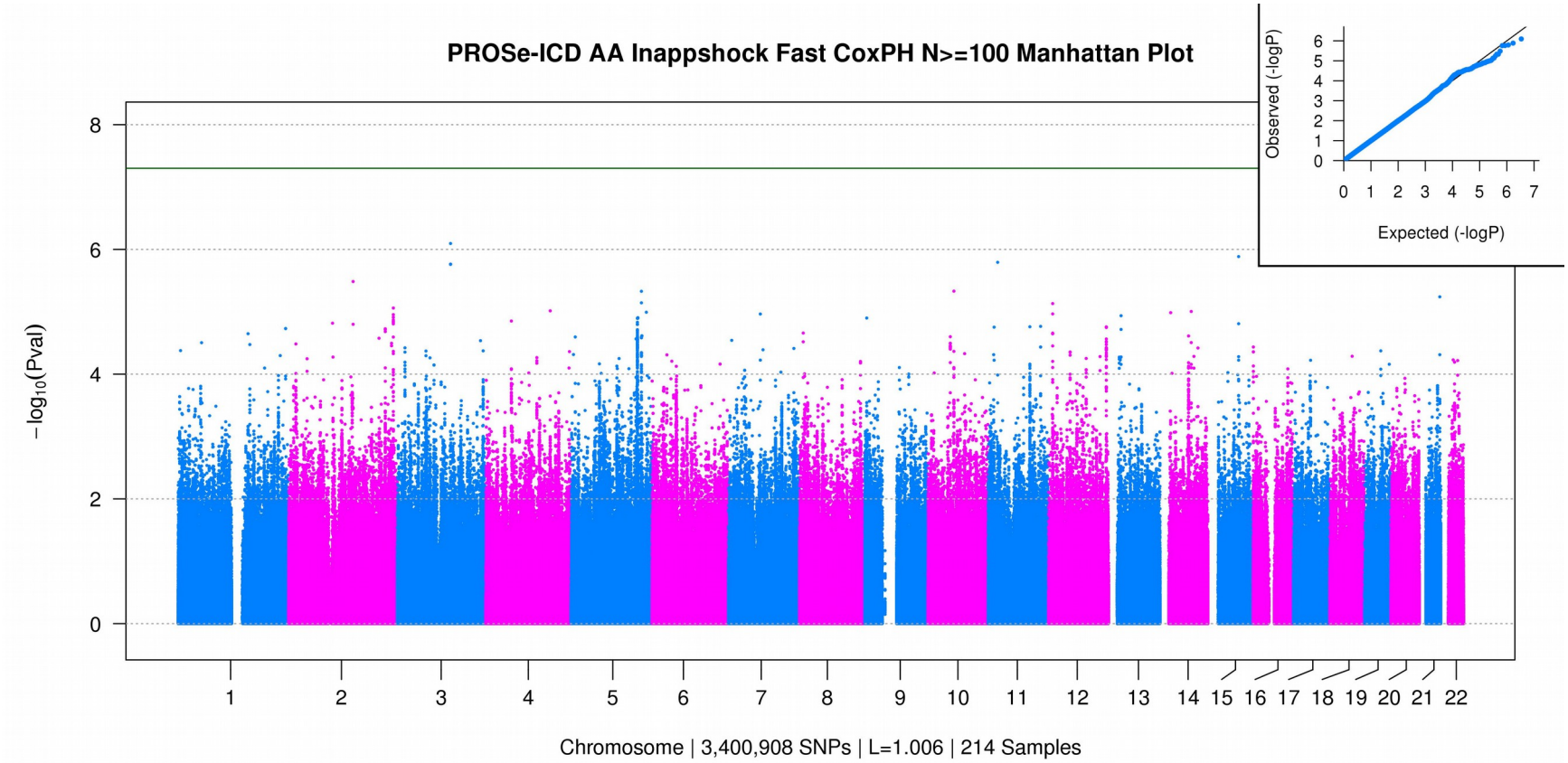

PROSe-ICD AA Inappshock Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 3,400,908 SNPs | L=1.006 | 214 Samples

Figure 4.6: PROSe-ICD AA Inapptherapy FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of inappropriate ICD therapy. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

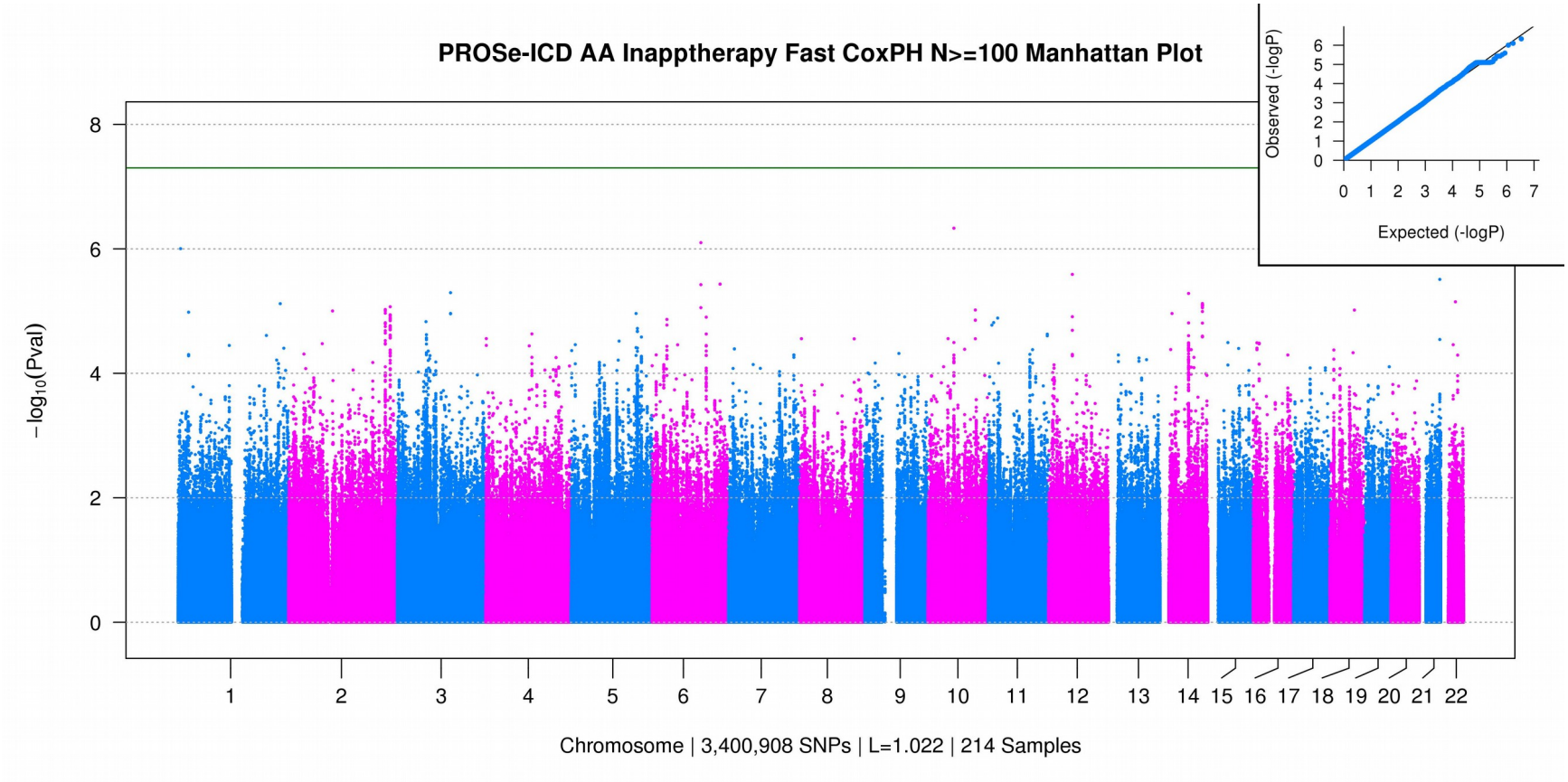

PROSe-ICD AA Inapptherapy Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
$-\log_{10}$(Pval)
Chromosome | 3,400,908 SNPs | L=1.022 | 214 Samples

Figure 4.7: PROSe-ICD AA QTc12 FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with African Americans from the PROSe-ICD cohort and the outcome phenotype of 12-lead Bazett corrected QT interval. Linear association model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper left. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

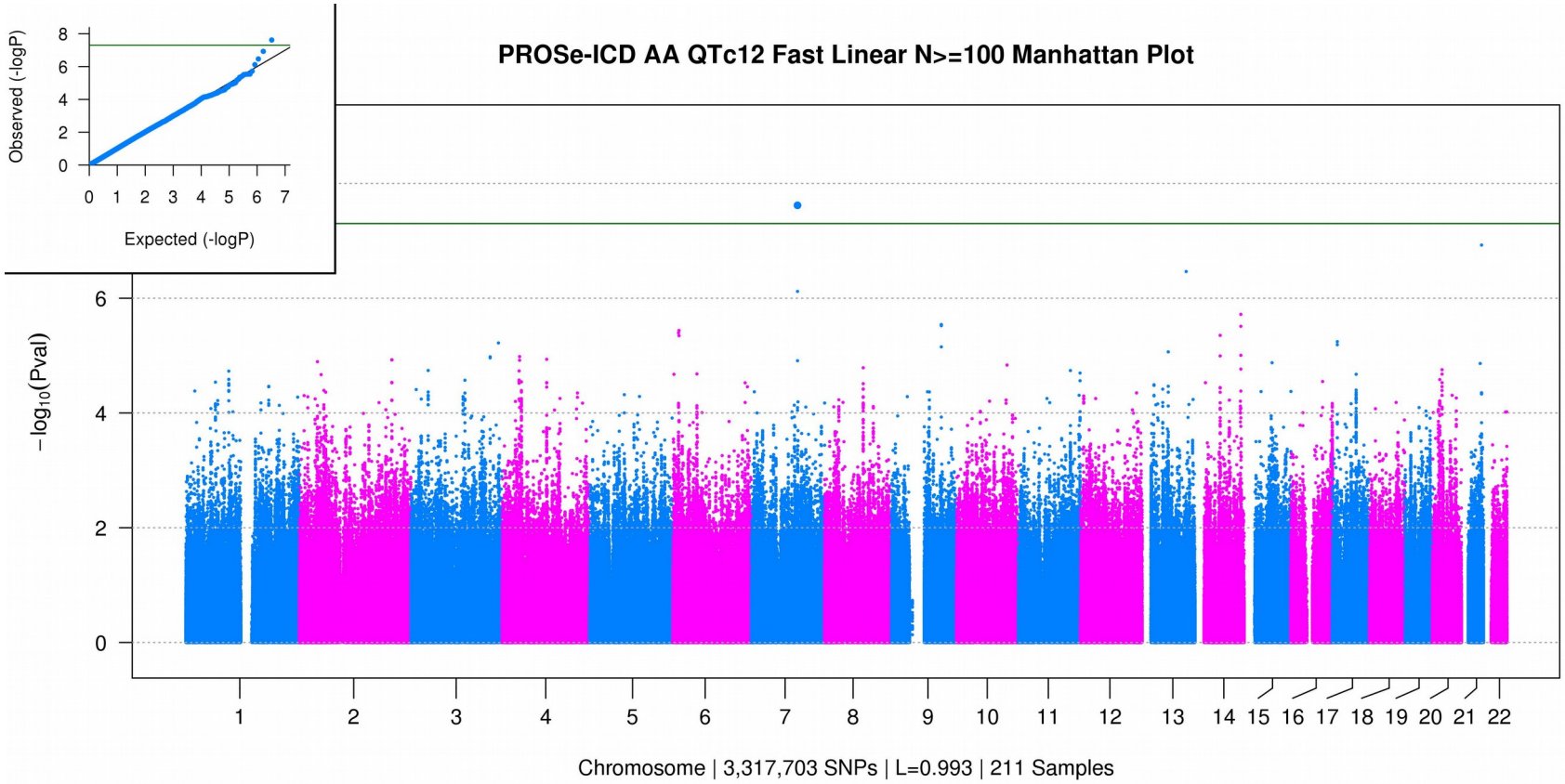

PROSe-ICD AA QTc12 Fast Linear N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
0 1 2 3 4 5 6 7 8
−log10(Pval)
0 2 4 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome | 3,317,703 SNPs | L=0.993 | 211 Samples

Figure 4.8: PROSe-ICD EA Appshock FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of appropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper left. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

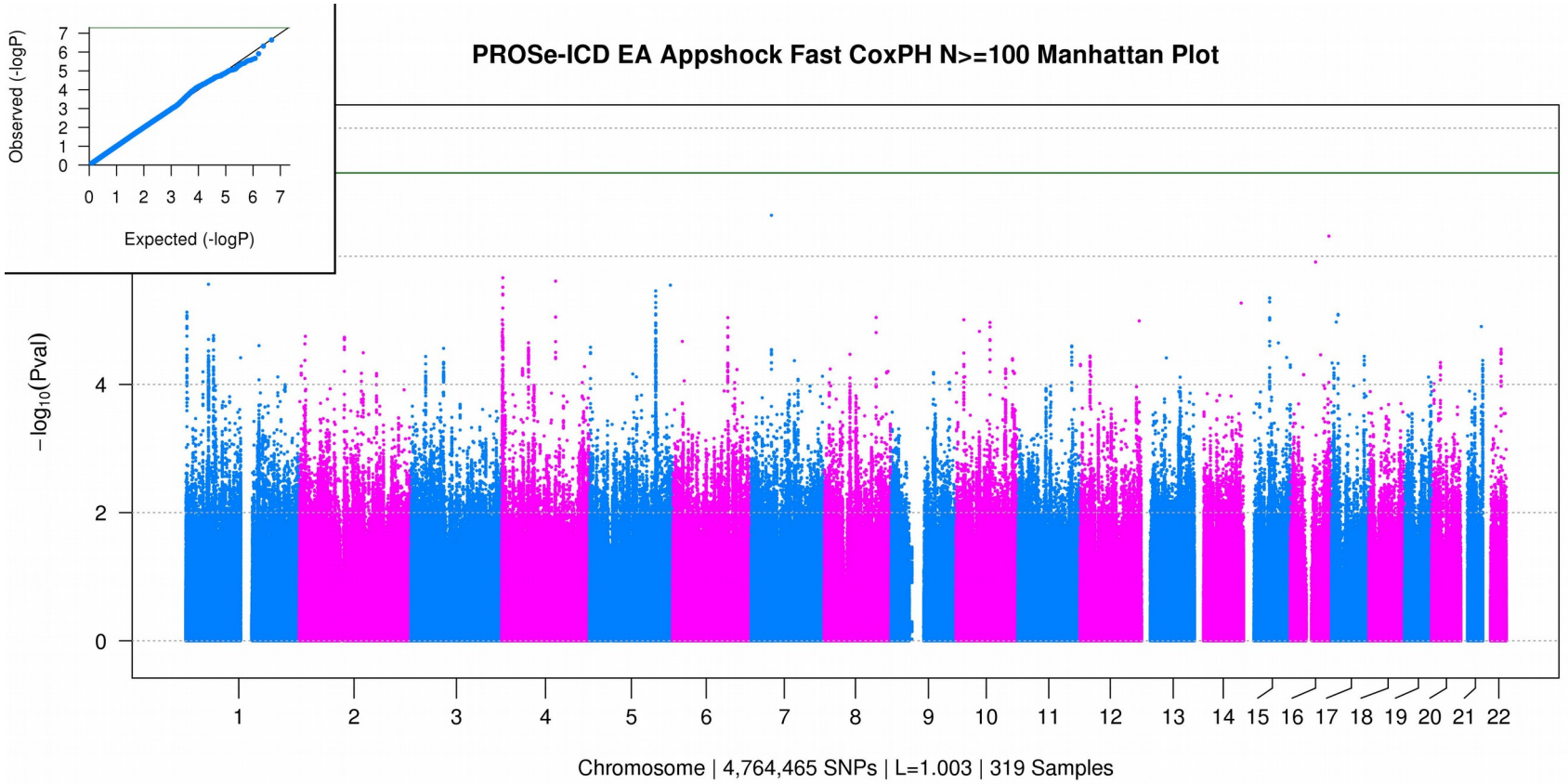

PROSe-ICD EA Appshock Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
0 1 2 3 4 5 6 7
$-\log_{10}$(Pval)
0 2 4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome | 4,764,465 SNPs | L=1.003 | 319 Samples

Figure 4.9: PROSe-ICD EA Apptherapy FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of appropriate ICD therapy. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

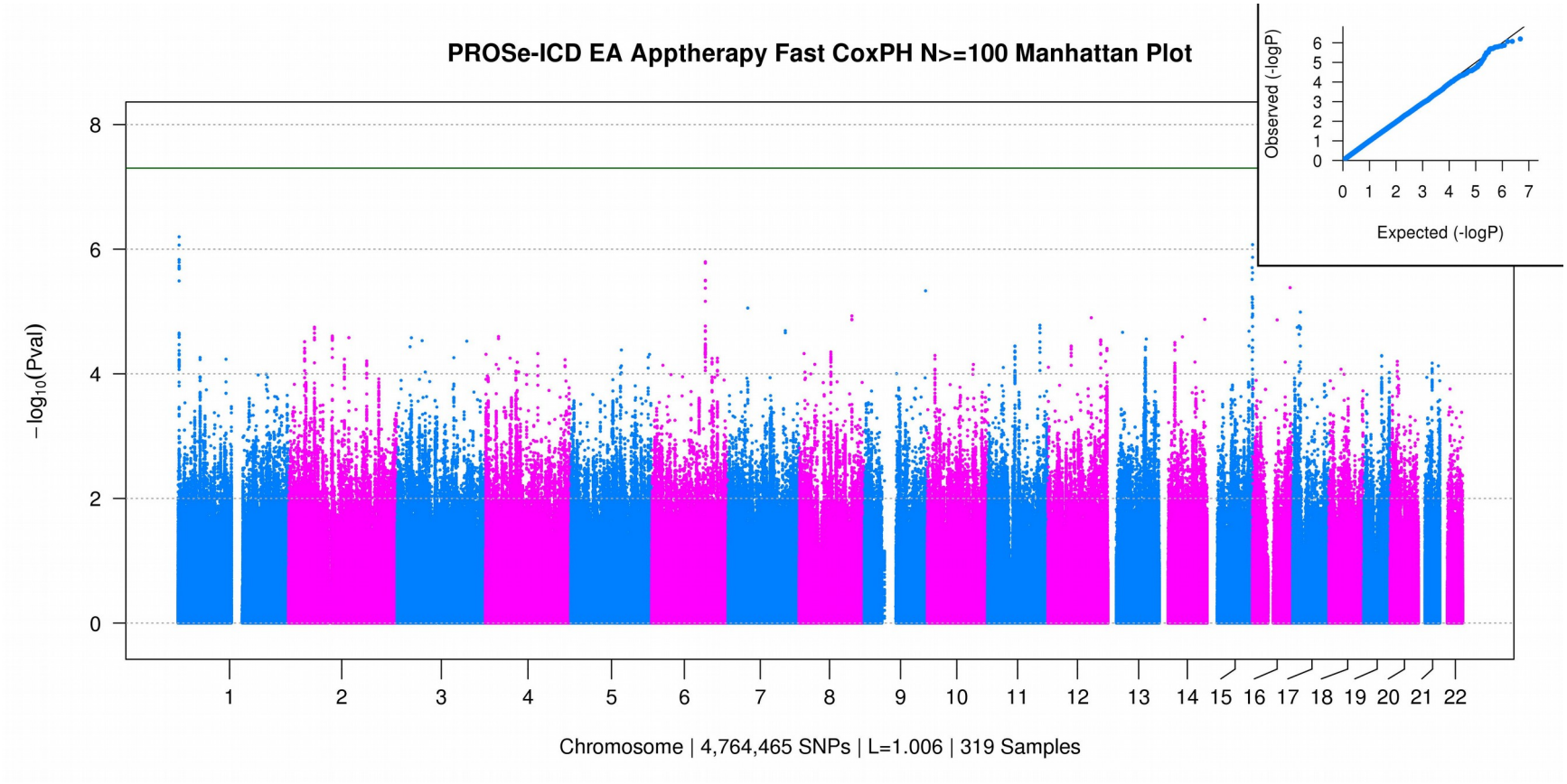

PROSe-ICD EA Apptherapy Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
$-\log_{10}$(Pval)
Chromosome | 4,764,465 SNPs | L=1.006 | 319 Samples

Figure 4.10: PROSe-ICD EA Death2 FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of all-cause mortality, censored at first appropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

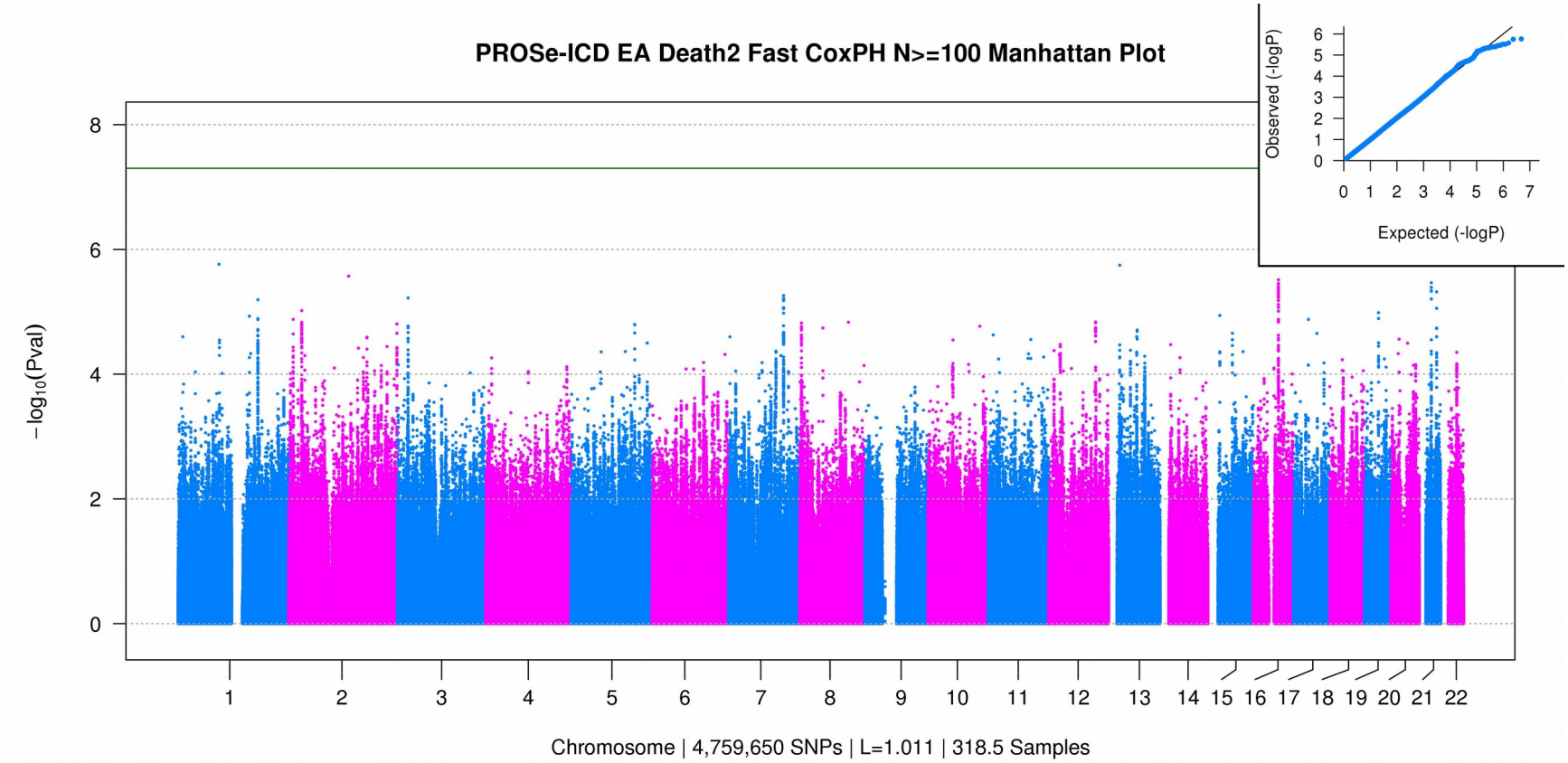

PROSe-ICD EA Death2 Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 4,759,650 SNPs | L=1.011 | 318.5 Samples

Figure 4.11: PROSe-ICD EA Death FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of all-cause mortality. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. "L" is the genomic inflation factor, lambda. "Samples" is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

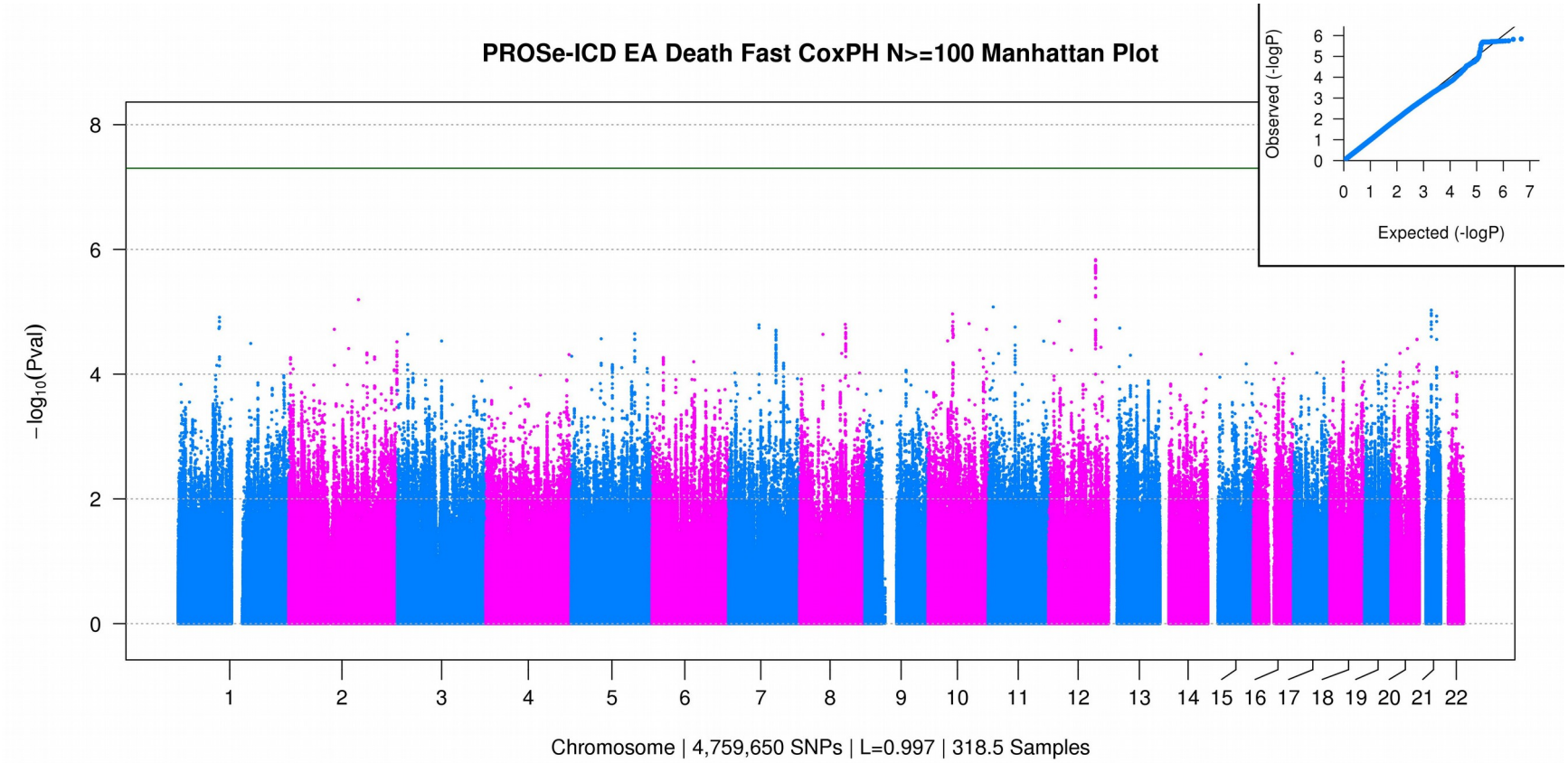

PROSe-ICD EA Death Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 4,759,650 SNPs | L=0.997 | 318.5 Samples

Figure 4.12: PROSe-ICD EA Inappshock FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of inappropriate ICD shock. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

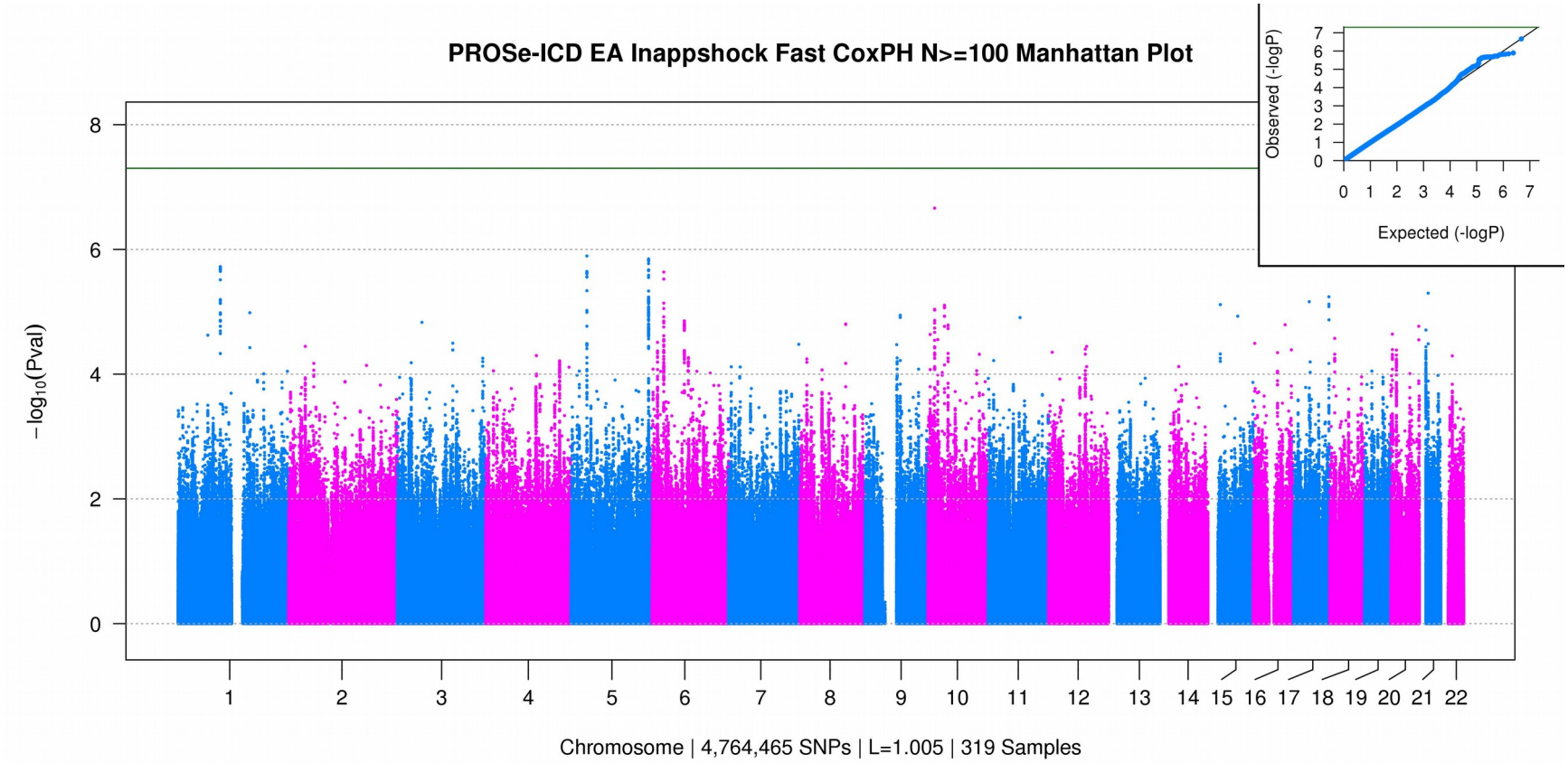

PROSe-ICD EA Inappshock Fast CoxPH N>=100 Manhattan Plot
−log10(Pval)
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome | 4,764,465 SNPs | L=1.005 | 319 Samples
Observed (-logP)
Expected (-logP)
0
1
2
3
4
5
6
7

Figure 4.13: PROSe-ICD EA Inapptherapy FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of inappropriate ICD therapy. CoxPH model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.

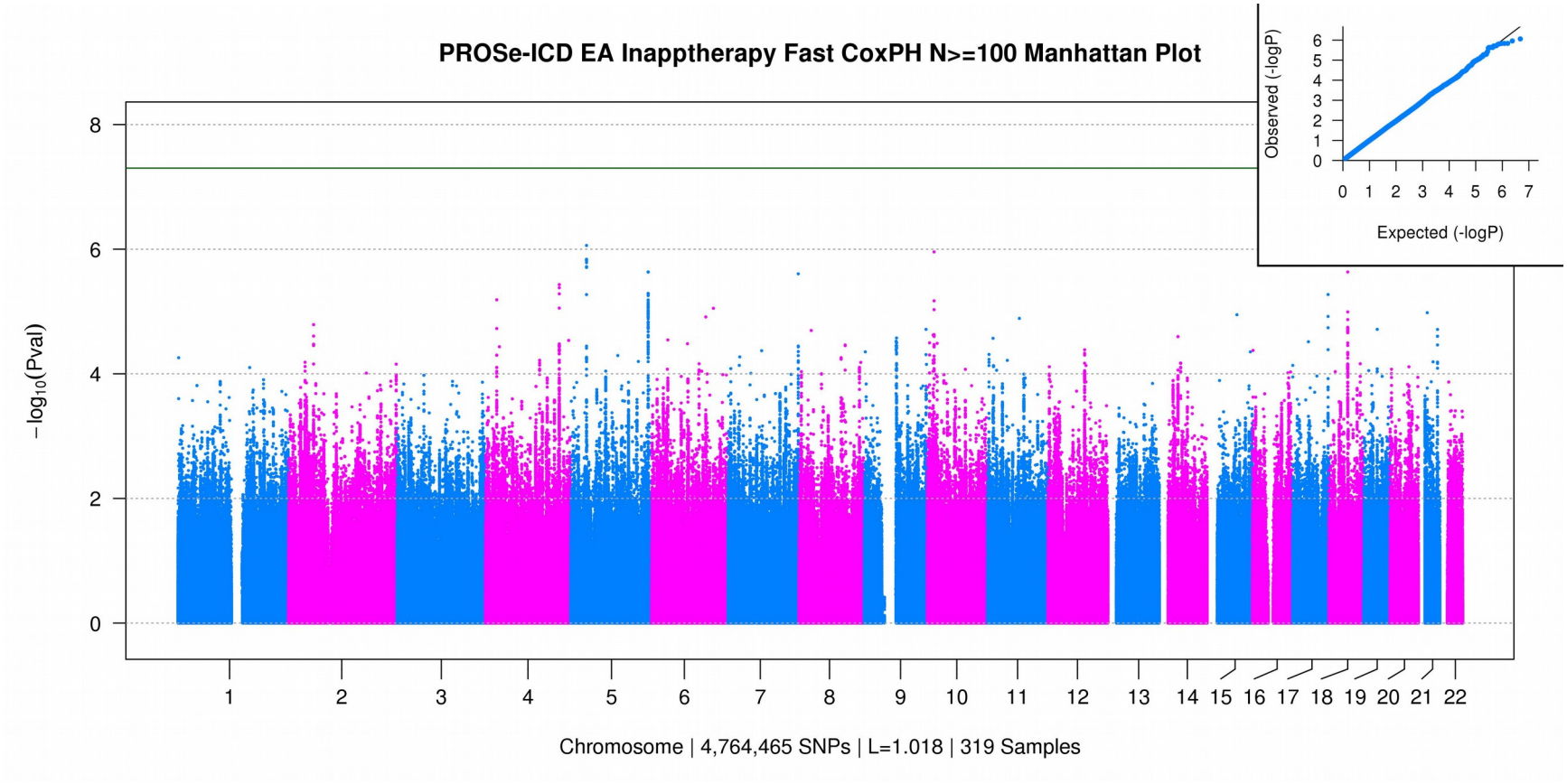
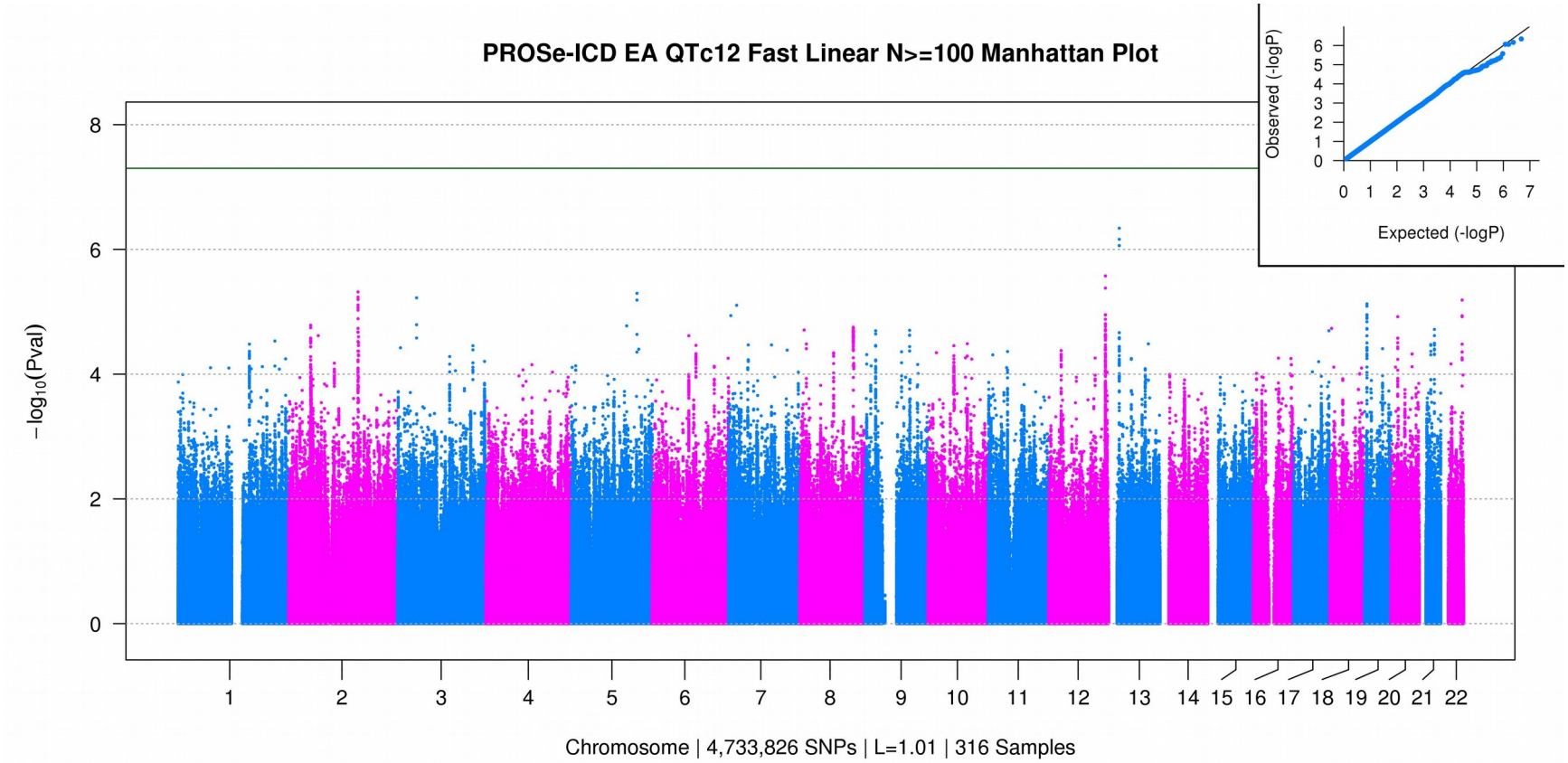

PROSe-ICD EA Inapptherapy Fast CoxPH N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 4,764,465 SNPs | L=1.018 | 319 Samples

Figure 4.14: PROSe-ICD EA QTc12 FAST CoxPH N>=100 Manhattan Plot

Manhattan plot of GWAS results with Caucasians from the PROSe-ICD cohort and the outcome phenotype of 12-lead Bazett corrected QT interval. Linear association model run using the FAST software. Results limited to effective sample size greater or equal to 100 individuals. X-axis is the human genome, autosomes-only. Y-axis is the significance level, -log10 of the p-value. The green bar represents the Bonferroni corrected significance level, p-value < 0.05/1000000. “L” is the genomic inflation factor, lambda. “Samples” is the maximum effective sample size calculated in FAST. Quantile-quantile plot of results in Manhattan plot is inset into upper right. X-axis represents expected significance under the null hypothesis. Y-axis is observed significance ordered by -log10 p-value. Early departure from the 45-degree line in black is indicative of an inflated test statistic, not seen here.


PROSe-ICD EA QTc12 Fast Linear N>=100 Manhattan Plot
Observed (-logP)
Expected (-logP)
−log10(Pval)
Chromosome | 4,733,826 SNPs | L=1.01 | 316 Samples

4.6 Conclusions

The only variant to reach genome-wide significance is rs148676350 in the African American QTc association. The locus has never been associated with QT interval in either the QT-IGC study or the QT interval study discussed in Chapter 3, although the variant itself was in neither study. rs148676350 is located in the 3' UTR of ZNF3 and has a coded allele frequency of 0.24, a Quality score of 0.71, an effective sample size of 109.6, an effect size estimate of 25.85, and a p-value of 2.40x10^-8.

We performed a power calculation to determine what we are likely to find using this study. We are well-powered to detect hazard ratios over 2.3 for SNPs with minor allele frequencies greater than 10%. If we had 10 times the sample size, we could detect hazard ratios over 1.3. Previously published genome-wide association studies indicate finding a common variant with hazard ratios that high is unlikely. Therefore, our results are not unexpected. We simply lack power to detect a genetic effect. To remedy this, we are working on building a consortium to increase sample size. We have started working with CardioDX's DISCERN cohort to this end and wish to build an ICD consortium.

Chapter 5: Conclusions

I have discussed three projects I undertook in the field of human genetics during the course of my graduate studies. The first was an investigation of the impact of genome-wide heterozygosity on overall longevity in humans. We consider "longevity" to be the lower risk of death within a time period. We found evidence that the protective effect of increased heterozygosity seen in lower organisms is present in humans. We estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death by within a given time period 1.57%. Our data shows this to be true even if the population itself has reduced mean heterozygosity. We observed a broad positive impact of genomic diversity on human survival demonstrated by the consistency we observed between European and African ancestry, males and females, and major causes of death. This project was concerned with human genetics at a population level. It is not a goal of this study to dissect the association to locate the underlying genetic and biologic causes or identify genes and pathways that contribute to the association.

The second project investigated of the role of coding variants in regulating QT Interval, a predictor of SCD. Our approach of focusing on coding variants identified 10 novel loci associated with QT/JT intervals, and highlighted the role of 17 specific genes, 7 of which were from novel loci. In addition to previously implicated pathways of potassium ion

regulation, sodium ion regulation, calcium ion regulation, and autonomic control of QT interval, our analyses highlighted a role for the internal structure of myocytes and interconnection of myocytes in modulating QT interval duration. We called this regulation mechanical control of QT interval. This project focused on a particular disease, SCD, by looking at one of its related phenotypes, QT interval, and applies human genetics to understand an individual's risk for this disease. The goal was finding individual genes that control electrophysiology and by extension that effect a person's risk for SCD and we found 17 genes effecting QT/JT intervals. This represents a narrowing of focus compared to the population-based question in the first project. These new discoveries will likely allow for the development of novel vectors for the prevention of lethal ventricular arrhythmias and SCD.

The last project was an investigation of the utility of genetics in predicting ICD therapy incidence. An accurate prediction can be used to evaluate if a particular patient should undergo surgery to receive an ICD. Unfortunately, our sample size was too limited to find meaningful results. We performed a power calculation to determine what we are likely to find using data from the PROSe-ICD cohort and found that we are well powered to detect hazard ratios over 2.3 in SNPs with minor allele frequencies greater than 10%. Judging by previously published genome-wide association studies, finding a common variant with hazard

ratios that high is unlikely. Sample size drives discovery and ours was too small for the expected hazard ratios. To remedy this, we are working on building a consortium to increase sample size. This project was concerned with applying human genetics to a clinical question. Finding an answer to this question would have directly affected the actions taken by medical doctors with regard to the manner which they treated patients.

Appendix A: Heterozygosity Publication

Bihlmeyer et al. BMC Genetics (2014) 15:159
DOI 10.1186/s12863-014-0159-7

BMC Genetics

RESEARCH ARTICLE Open Access

Genetic diversity is a predictor of mortality in humans

Nathan A Bihlmeyer[1,2], Jennifer A Brody[35], Albert Vernon Smith[32,33], Kathryn L Lunetta[9,10], Mike Nalls[6], Jennifer A Smith[14], Toshiko Tanaka[36], Gail Davies[15,16], Lei Yu[18], Saira Saeed Mirza[21], Alexander Teumer[27,28], Josef Coresh[38], James S Pankow[39], Nora Franceschini[40], Anish Scaria[3], Junko Oshima[4], Bruce M Psaty[5], Vilmundur Gudnason[32,33], Gudny Eiriksdottir[32], Tamara B Harris[34], Hanyue Li[9], David Karasik[12], Douglas P Kiel[12], Melissa Garcia[7], Yongmei Liu[8], Jessica D Faul[13], Sharon LR Kardia[14], Wei Zhao[14], Luigi Ferrucci[36], Michael Allerhand[15], David C Liewald[15], Paul Redmond[16], John M Starr[15,17], Philip L De Jager[19], Denis A Evans[20], Nese Direk[21], Mohammed Arfan Ikram[21,22,23], André Uitterlinden[21,26], Georg Homuth[27], Roberto Lorbeer[28], Hans J Grabe[29,30], Lenore Launer[34], Joanne M Murabito[10,11], Andrew B Singleton[6], David R Weir[13], Stefania Bandinelli[37], Ian J Deary[15,16], David A Bennett[18], Henning Tiemeier[21,24,25], Thomas Kocher[31], Thomas Lumley[3*] and Dan E Arking[2*]

Abstract

Background: It has been well-established, both by population genetics theory and direct observation in many organisms, that increased genetic diversity provides a survival advantage. However, given the limitations of both sample size and genome-wide metrics, this hypothesis has not been comprehensively tested in human populations. Moreover, the presence of numerous segregating small effect alleles that influence traits that directly impact health directly raises the question as to whether global measures of genomic variation are themselves associated with human health and disease.

Results: We performed a meta-analysis of 17 cohorts followed prospectively, with a combined sample size of 46,716 individuals, including a total of 15,234 deaths. We find a significant association between increased heterozygosity and survival ($P = 0.03$). We estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death by 1.57%.

Conclusions: This effect was consistent between European and African ancestry cohorts, men and women, and major causes of death (cancer and cardiovascular disease), demonstrating the broad positive impact of genomic diversity on human survival.

Keywords: Heterozygosity, Human, Survival, GWAS

* Correspondence: t.lumley@auckland.ac.nz; arking@jhmi.edu
[3]Department of Statistics, University of Auckland, 303.325 Science Centre, Private Bag 92019, Auckland 1142, New Zealand
[2]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, BRB Room 447, 733 N. Broadway St, Baltimore, MD 21205, USA
Full list of author information is available at the end of the article

Background

With the advent of genome-wide association studies (GWAS), and more recently whole-exome and whole-genome sequencing, remarkable progress has been made in elucidating the genetics of complex traits, with numerous genetic variants each explaining a small fraction of the variance [1,2]. The presence of numerous segregating small effect alleles within the genome that influence traits that directly impact health raises the question of whether global measures of genomic variation are themselves associated with human health and disease. Indeed, increased fitness has been associated with the increase of genetic diversity across many organisms [3,4], including humans [5-8], and is often referred to as positive Heterozygosity Fitness Correlations (HFCs). In particular, associations have been found between heterozygosity at the Major Histocompatibility Complex (MHC) (a.k.a. Human Leukocyte Antigen, HLA) region and general health in humans [9]. In the case of heterozygosity in the MHC region, the cause of a positive HFC being observed is believed to be the result of increased antibody diversity conveying robust pathogen resistance and therefore increased general health [10]. However, in the case of increased whole-genome heterozygosity, the mechanism of action is less readily apparent. Two general mechanisms that act at a genome level to influence fitness have been proposed. The first is compensation for recessive deleterious mutations [11], whereas the second is a specific advantage of the heterozygous state over either homozygous state (overdominance/heterozygous advantage) [11], such as that observed for the sickle cell mutation in the presence of endemic malarial disease. It has been proposed that compensation for deleterious mutations occurs at many loci and is the major mechanism at work in HFCs, with overdominance occurring at few loci but with greater effect size per occurrence [11].

Results and discussion

Various heterozygosity metrics have been proposed [12]. The heterozygosity metric used in this study is the sum of all heterozygous loci divided by the expected state given the allele frequency under Hardy-Weinberg Equilibrium $t = \frac{\sum_{0;1}}{\sum 2p(1-p)}$: where p is the frequency of the major allele in each cohort. This metric up-weights loci where the expectation of being heterozygous is low. Given the relationship between effect size and allele frequency [13,14], up-weighting loci with low minor allele frequencies should maximize the ability to detect a HFC in humans under a model in which the compensation for deleterious alleles is the major mechanism driving HFCs. Only Single Nucleotide Polymorphisms (SNPs) on the autosomes were considered.

To test for the effect of genome-wide heterozygosity on survival, we performed a meta-analysis of 17 cohorts (13 European ancestry, 4 African American ancestry) followed prospectively, with a combined sample size of 46,716 individuals, including a total of 15,234 deaths (Additional file 1: Table S1). Within each cohort, a Cox proportional hazards model (CoxPH) was used comparing age at study entry to age at study exit (death) or most recent follow-up (alive), and included covariates known to affect survival (sex, highest education level, Body Mass Index (BMI), income level, center where DNA was collected, and the first ten principal components to adjust for population substructure). Since each cohort used a different number of SNPs (Additional file 1: Table S1), the variances of the heterozygosity metrics are not the same (they are dependent on the total number of SNPs in the metric), and effect sizes from each cohort are not directly comparable. Using Stouffer's method to combine Z-scores, weighted by the number of deaths in each cohort, we find a significant association between increased heterozygosity and survival (P = 0.03). To assess effect size, we standardized the beta estimates by multiplying them by the standard deviation of the heterozygosity metric for each cohort [15]. This method does not completely account for the aforementioned bias; however, it is the most appropriate method to determine an interpretable effect size. Combining the standardized beta estimates using inverse variance weighting demonstrates that for every standard deviation increase in heterozygosity a person has over the population mean, they are expected to have a 1.57% decreased risk of death (Figure 1). There was no evidence for heterogeneity across studies, and a direct comparison of European Ancestry to African ancestry cohorts showed no significant difference (Figure 2, P = 0.80); thus, all downstream analyses combined European and African ancestry cohorts.

To test whether all chromosomes are contributing equally to the association between heterozygosity and survival, each study subject's heterozygosity score was recalculated using only SNPs from a given chromosome. An inverse-variance meta-analysis for each chromosome was performed across studies, followed by a meta-analysis of the chromosomal results (Figure 3). No significant difference was observed between effects across chromosomes (P = 0.17). To test whether all major causes of death contribute equally to our genome-wide finding, death caused by cancer, death caused by CVD, and other causes of death were each analyzed separately. A meta-analysis for each cause of death was performed as described above, followed by a test for heterogeneity and model fitting. Our results demonstrate that heterozygosity is protective for all causes of death, with no significant evidence for heterogeneity (Figure 4, P = 0.79). To assess if heterozygosity levels impact women differently from men, meta-analyses

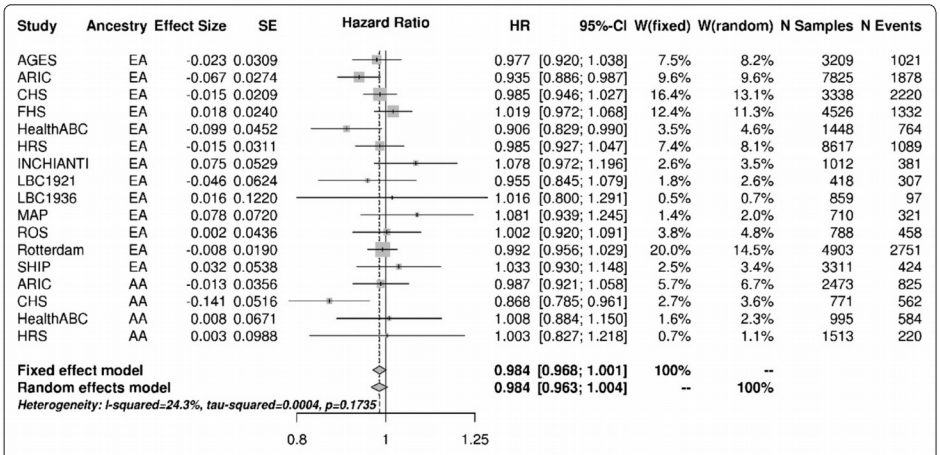


Figure 1 Heterozygosity meta-analysis by study. 1.57% decreased risk of death for every standard deviation increase in heterozygosity. This is determined using an inverse variance weighted fixed effect model. Significance of P = 0.03 is determined using Stouffer's method to combine Z-scores due to bias in inverse variance weighted fixed effect model. There are 46,716 individuals, including a total of 15,234 deaths. EA = European Ancestry; AA = African Ancestry; AGES = Age, Gene/Environment Susceptibility cohort; ARIC = Atherosclerosis Risk In Communities cohort; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; HealthABC = HealthABC cohort; HRS = Health and Retirement Study; INCHINTI = InCHIANTI cohort; LBC1921 = 1921 Lothian Birth Cohort; LBC1936 = 1936 Lothian Birth Cohort; MAP = Rush Memory and Aging Project cohort; ROS = Religious Orders Study; Rotterdam = Rotterdam Study; SHIP = Study of Health In Pomerania cohort; SE = Standard Error; HR = Hazard Ratio; CI = Confidence Interval; W = Weight; N = Number.

were performed separately for each sex. Our results do not provide evidence for a differential effect of heterozygosity on survival in men vs. women (Figure 5, P = 0.49).

Conclusions

In summary, this study provides evidence that the protective effect of increased heterozygosity seen in lower organisms functions in humans as well and may have implications for how we design future studies to identify genetic determinants of human disease and survival. We estimate that within a single population, every standard deviation of heterozygosity an individual has over the mean decreases that person's risk of death by 1.57%. Interestingly, this seems to be true even if the population itself has reduced mean heterozygosity. In future studies, limiting to heterozygosity in proximity to genes and/or regulatory elements may reveal if some regions are more sensitive to heterozygosity than others. Increasing the African ancestry sample size may increase power to see a difference between ancestry groups. Overall the consistency we observed between European and African ancestry, males and females, and major causes of death demonstrate a broad positive impact of genomic diversity on human survival.

Methods

Methods for each individual cohort can be found in Additional file 2: Text S1. Self-described Caucasian ("white", "Caucasian") and African ancestry ("black", "African American") individuals were included after excluding first and second degree relatives and genetic outliers. Genetic outliers were defined by merging genotyping data with HapMap3 data, and calculating the Euclidean distance from a combined reference HapMap3 population

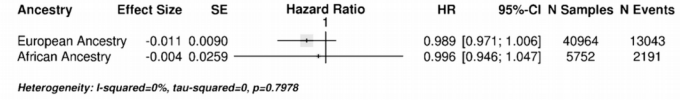


Figure 2 Ancestry meta-analysis. Direct comparison of European Ancestry to African ancestry cohorts showed no significant difference (P = 0.80). Figure is formatted the same as Figure 1.

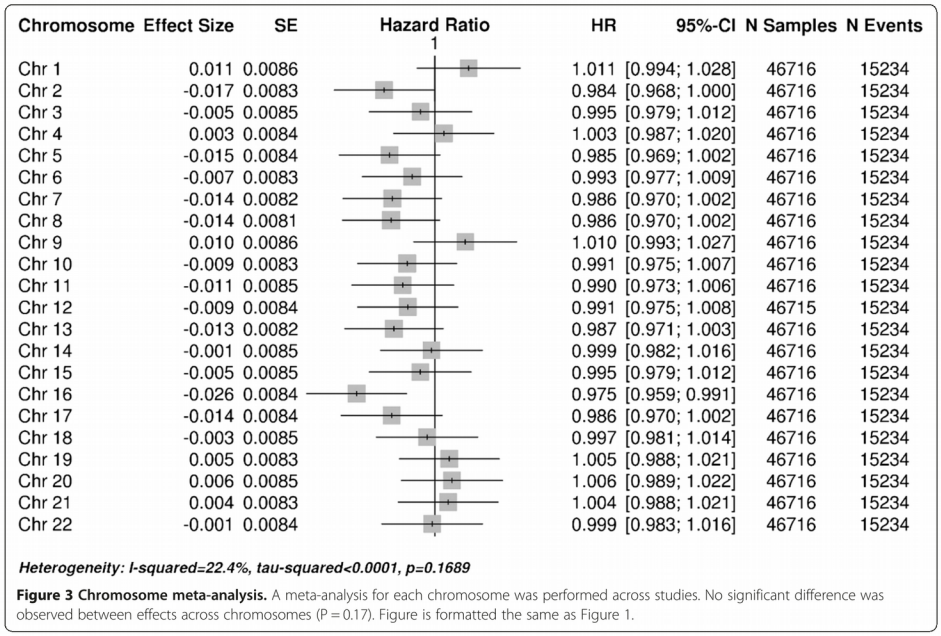


Figure 3 Chromosome meta-analysis. A meta-analysis for each chromosome was performed across studies. No significant difference was observed between effects across chromosomes (P = 0.17). Figure is formatted the same as Figure 1.

(Caucasian = CEU + TSI, African ancestry = ASW + YRI + MKK + LWK) cluster centroid in the first 3 PC space weighted by explained variance. Specifically, the standard deviation of Euclidean distance was determined for each HapMap reference group, and any sample greater than ten standard deviations away from centroid were defined as genetic outliers and excluded.

Directly genotyped SNPs were used for all analyses (Additional file 3: Figure S1). Imputed SNPs were not used to avoid issues with genotype accuracy and bias towards the reference panel. SNP exclusion criteria included: monomorphic in the dataset, non-unique mapping to Hg19, SNPs which are no longer in the company provided annotation file for the SNP array, >0.5% missing data, MAF ≤ 10%, HWE p-value ≥ 0.001, and non-autosomal SNPs. The heterozygosity metric is the sum of all heterozygous loci divided by the expected state given the allele frequency under Hardy-Weinberg Equilibrium: $t = \frac{\sum_{0;1}}{\sum 2p(1-p)}$ where p is the frequency of the major allele.

Separate association analyses were run for Caucasian and African ancestry samples from each cohort. The Cox Proportional Hazard Model (CoxPH) included covariates for Body Mass Index (BMI) at first visit and first ten principal components, and the 'strata' function for sex, education level (defined as 1. ≤11th grade, 2. high school diploma, general equivalence diploma or some vocational school, 3. 1–4 years of college, 4. Some graduate/professional school, and Missing), income level (defined by cohorts), and center of DNA collection within cohorts. The CoxPH model was set up so that the outcome was age at study entry, age at study exit, and a binary variable coding state of death (1: Dead, 0: Alive). Age is measured in units of years, but is accurate to the nearest day.

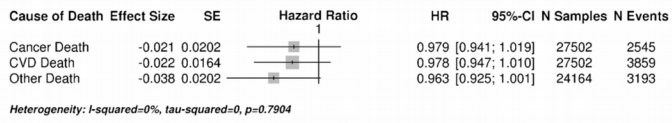


Figure 4 Causes of death meta-analysis. A meta-analysis for each cause of death was performed. Our results show no significant evidence for heterogeneity (Figure 4, P = 0.79). Figure is formatted the same as Figure 1.

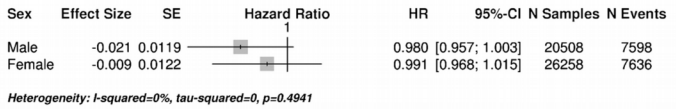


Figure 5 Sex meta-analysis. A meta-analysis was performed separately for each sex. Our results do not provide evidence for a differential effect of heterozygosity on survival in men vs. women (Figure 5, P = 0.49). Figure is formatted the same as Figure 1.

For the meta-analysis, significance was determined by Stouffer's method [16] calculated as a two-sided test by incorporating Z-scores derived from two-sided tests performed in each cohort. We standardized the beta estimates by multiplying them by the standard deviation of the heterozygosity metric for each cohort, to account for the fact that the effect size is proportional to the variance in the heterozygosity metric. The variance heterozygosity metric in turn is proportional to the inverse of the square root of the number of SNPs used to determine the heterozygosity metric. Because most cohorts used different genotyping arrays, a large bias is introduced into the meta-analysis. Stouffer's method completely removes this bias; however, cannot estimate a combined effect size, only the overall significance. To get an estimate of the combined effect size (recognizing that the P-value and associated confidence intervals will be inflated), we used inverse variance weighting of the standardized cohort effect sizes, which partially corrects the bias and allows for the combined effect size to be estimated.

Ethics statements

Institutional Review Board approvals were obtained by each participating ARIC study center (the Universities of NC, MS, MN, and John Hopkins University) and the coordinating center (University of NC), and the research was conducted in accordance with the principles described in the Helsinki Declaration. All subjects in the ARIC study gave informed consent. For more information see dbGaP Study Accession: phs000280.v2.p1. JHSPH IRB number H.34.99.07.02.A1. Manuscript proposal number MS1964.

HealthABC Human subjects protocol UCSF IRB is H5254-12688-11.

CHS was approved by institutional review committees at each site, the subjects gave informed consent, and those included in the present analysis consented to the use of their genetic information for the study of cardiovascular disease. It is the position of the UW IRB that these studies of de-identified data, with no patient contact, do not constitute human subjects research. Therefore we have neither an approval number, nor an exemption.

IRB permission to conduct genetics-related work in the Health and Retirement Study (HRS) is granted under the project title, "Expanding a National Resource for Genetic Research in Behavioral & Health Science" (HUM00063444). The IRB that approved this project is the Health Sciences and Behavioral Sciences Institutional Review Board at the University of Michigan. No manuscript proposal is required for use of HRS data.

Inchianti ethics review statement: The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD).

The Religious Orders Study (ORA# 91020181) and the Rush Memory and Aging Project (ORA# 86121802) were approved by the Institutional Review Board of Rush University Medical Center. Written informed consent was obtained from all the participants.

The SHIP study followed the recommendations of the Declaration of Helsinki. The study protocol of SHIP was approved by the medical ethics committee of the University of Greifswald. Written informed consent was obtained from each of the study participants. The SHIP study is described in PMID: 20167617.

The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. A written informed consent was obtained from all participants.

The Boston University Medical Campus Institutional Review Board approved the FHS genome-wide genotyping (protocol number H-226671) and genetic investigation of aging and longevity phenotypes (protocol number H-24912).

The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, (VSN: 00–063) and the Data Protection Authority. The researchers are indebted to the participants for their willingness to participate in the study.

Ethics permission for the LBC studies was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LBC1936: LREC/2003/2/29 and LB

C1921: LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent.

Additional files

Additional file 1: Table S1. Descriptive breakdown of each cohort and summary statistics.

Additional file 2: Text S1. Additional Methods for each individual cohort.

Additional file 3: Figure S1. Heterozygosity Metrics Determined Using Different SNP Lists. The dataset used was genome wide SNP data from sequencing of 503 individuals with European ancestry from 1000G phase 3 release. The SNP lists used were: 1) all SNPs 2) SNPs on the Illumina 1M 3) SNPs on the Illumina 610quad 4) SNPs on the Illumina Omni2.5 and 5) SNPs on the Affymetrix 6.0. This is to determine if SNP selection on the arrays biases the heterozygosity metric. We see high correlation and no systematic bias.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Designed Study: NAB, TL, and DEA. Ran Analyses: NAB, JAB, AVS, KLL, MN, JAS, TT, GD, LY, SSM, AT. Contributed Data: JC, JSP, NF, AS, JO, BMP, VG, GE, TBH, HL, DK, DPK, MG, YL, JDF, SLRK, WZ, LF, MA, DCL, PR, JMS, PLD, DAE, ND, MAI, AU, GH, RL, HJG, LL, JMM, ABS, DRW, SB, IJD, DAB, HT, TK, TL, DEA. All authors read and approved the final manuscript.

Acknowledgements
Funding
Funded in part by training grant (NIGMS) 5T32GM07814.
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1232825. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

Cohorts
ARIC
The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.
AGES
The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, (VSN: 00–063) and the Data Protection Authority. The researchers are indebted to the participants for their willingness to participate in the study.
CHS
Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756, HL085251 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org/.
The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
FHS
Funding: The Framingham Heart Study analyses were supported by the National Institute of Aging (R01AG29451). This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Dr. Kiel was partially supported by the National Institute of Arthritis Musculoskeletal and Skin Diseases (R01 AR41398).
HealthABC
This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106 and was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (Z01 AG000949-02 and Z01 AG007390-07, Human subjects protocol UCSF IRB is H5254-12688-11). The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov).
HRS
HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.
InCHIANTI
The InCHIANTI study baseline (1998–2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).
LBC
Lothian Birth Cohorts 1921 and 1936 (LBC1921, LBC1936)
We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the BBSRC, The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping of the cohorts was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC, and Medical Research Council (MRC) is gratefully acknowledged.
MAP/ROS
The MAP and ROS data used in this analysis was supported by National Institute on Aging grants P30AG10161, R01AG17917, R01AG15819, R01AG30146, the Illinois Department of Public Health, and the Translational Genomics Research Institute.
Rotterdam
The Rotterdam Study is supported by Erasmus Medical Centre and Erasmus University Rotterdam, the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the

Netherlands Genomics Initiative, the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.
Prof. Tiemeier was supported by the VIDI grant of ZonMw (2009–017.106.370). Dr. Ikram was supported by the VENI grant of NWO. The funders had no role in the study design or data collection and analysis.

SHIP

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

Author details

[1]Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, BRB Room 447, 733 N. Broadway St, Baltimore, MD 21205, USA. [3]Department of Statistics, University of Auckland, 303.325 Science Centre, Private Bag 92019, Auckland 1142, New Zealand. [4]Department of Pathology, University of Washington, Seattle, WA, USA. [5]Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA. [6]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [7]Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [8]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA. [9]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [10]The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA. [11]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [12]Institute for Aging Research, Hebrew Senior Life, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Cambridge, MA, USA. [13]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [14]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [15]Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK. [16]Department of Psychology, The University of Edinburgh, Edinburgh, UK. [17]Alzheimer Scotland Dementia Research Centre, The University of Edinburgh, Edinburgh, UK. [18]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [19]Program in Translational NeuroPsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [20]Rush Institute for Healthy Aging and Department of Internal Medicine, Rush University Medical Center, Chicago, IL, USA. [21]Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands. [22]Department of Neurology, Erasmus Medical Centre, Rotterdam, The Netherlands. [23]Department of Radiology, Erasmus Medical Centre, Rotterdam, The Netherlands. [24]Department of Child and Adolescent Psychiatry, Erasmus Medical Centre, Rotterdam, The Netherlands. [25]Department of Psychiatry, Erasmus Medical Centre, Rotterdam, The Netherlands. [26]Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands. [27]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [28]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [29]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, HELIOS Hospital Stralsund, Greifswald, Germany. [30]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Greifswald, Germany. [31]Unit of Periodontology, Department of Restorative Dentistry, Periodontology and Endodontology, University Medicine Greifswald, Greifswald, Germany. [32]Icelandic Heart Association, Kopavogur, Iceland. [33]University of Iceland, Reykjavik, Iceland. [34]National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [35]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [36]Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA. [37]Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy. [38]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [39]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA. [40]Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA.

Received: 22 July 2014 Accepted: 19 December 2014
Published online: 29 December 2014

References

1. **Online Mendelian Inheritance in Man** [http://www.omim.org/]
2. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci U S A* 2009, **106:**9362–9367.
3. Mitton JB, Grant MC: **Associations among protein heterozygosity, growth rate, and developmental homeostasis.** *Annu Rev Ecol Syst* 1984, **15:**479–499.
4. Alibert P, Renaud S, Dod B, Bonhomme F, Auffray JC: **Fluctuating asymmetry in the Mus musculus hybrid zone: a heterotic effect in disrupted co-adapted genomes.** *Proc Biol Sci* 1994, **258:**53–59.
5. Roberts SC, Little AC, Gosling LM, Perrett DI, Carter V, Jones BC, Penton-Voak I, Petrie M: **MHC-heterozygosity and human facial attractiveness.** *Evol Hum Behav* 2005, **26:**213–226.
6. Coetzee V, Barrett L, Greeff JM, Henzi SP, Perrett DI, Wadee AA: **Common HLA alleles associated with health, but Not with facial attractiveness.** *PLoS One* 2007, **2:**e640.
7. Campbell H, Carothers AD, Rudan I, Hayward C, Biloglav Z, Barac L, Pericic M, Janicijevic B, Smolej-Narancic N, Polasek O, Kolcic I, Weber JL, Hastie ND, Rudan P, Wright AF: **Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits.** *Hum Mol Genet* 2007, **16:**233–241.
8. Takata H, Ishii T, Suzuki M, Sekiguchi S, Iri H: **Influence of major histocompatibility complex region genes on human longevity among okinawan-japanese centenarians and nonagenarians.** *Lancet* 1987, **330:**824–826.
9. Lie HC, Simmons LW, Rhodes G: **Does genetic diversity predict health in humans?** *PLoS One* 2009, **4:**e6391.
10. Piertney SB, Oliver MK: **The evolutionary ecology of the major histocompatibility complex.** *Heredity* 2005, **96:**7–21.
11. Charlesworth D, Willis JH: **The genetics of inbreeding depression.** *Nat Rev Genet* 2009, **10:**783–796.
12. Szulkin M, Bierne N, David P: **Heterozygosity-fitness correlations: a time for reappraisal.** *Evolution* 2010, **64:**1202–1217.
13. Arking DE, Chakravarti A: **Understanding cardiovascular disease through the lens of genome-wide association studies.** *Trends Genet TIG* 2009, **25:**387–394.
14. Hindorff LA, Gillanders EM, Manolio TA: **Genetic architecture of cancer and other complex diseases: lessons learned and future directions.** *Carcinogenesis* 2011, **32:**945–954.
15. Menard S: **Six approaches to calculating standardized logistic regression coefficients.** *Am Stat* 2004, **58:**218–223.
16. Stouffer Samuel A, Suchman Edward A, DeViney Leland C, Star Shirley A, Williams Robin M Jr: *The American Soldier. Adjusting During Army Life, Vol. 1.* Princeton: Princeton University Press; 1949.

References

Alibert, P et al. “Fluctuating Asymmetry in the Mus Musculus Hybrid Zone: A Heterotic Effect in Disrupted Co-Adapted Genomes.” *Proceedings. Biological sciences / The Royal Society* 258.1351 (1994): 53–59. *NCBI PubMed*. Web.

Anderson, Garnet L. et al. “Implementation of the Women’s Health Initiative Study Design.” *Annals of Epidemiology* 13.9 Suppl (2003): S5-17. Print.

Arking, Dan E., Arne Pfeufer, et al. “A Common Genetic Variant in the NOS1 Regulator NOS1AP Modulates Cardiac Repolarization.” *Nature Genetics* 38.6 (2006): 644–651. *PubMed*. Web.

Arking, Dan E., Sara L. Pulit, et al. “Genetic Association Study of QT Interval Highlights Role for Calcium Signaling Pathways in Myocardial Repolarization.” *Nature Genetics* 46.8 (2014): 826–836. *PubMed*. Web.

Arking, Dan E, and Aravinda Chakravarti. “Understanding Cardiovascular Disease through the Lens of Genome-Wide Association Studies.” *Trends in genetics: TIG* 25.9 (2009): 387–394. *NCBI PubMed*. Web.

Bennett, David A, Julie A Schneider, Zoe Arvanitakis, et al. “Overview and Findings from the Religious Orders Study.” *Current Alzheimer research* 9.6 (2012): 628–645. Print.

Bennett, David A, Julie A Schneider, Aron S Buchman, et al. “Overview and Findings from the Rush Memory and Aging Project.” *Current Alzheimer research* 9.6 (2012): 646–663. Print.

Bihlmeyer, Nathan A. et al. “Genetic Diversity Is a Predictor of Mortality in Humans.” *BMC Genetics* 15 (2014): 159. *BioMed Central*. Web.

Bild, Diane E. et al. “Multi-Ethnic Study of Atherosclerosis: Objectives and Design.” *American Journal of Epidemiology* 156.9 (2002): 871–881. Print.

Campbell, Harry et al. “Effects of Genome-Wide Heterozygosity on a Range of Biomedically Relevant Human Quantitative Traits.” *Human Molecular Genetics* 16.2 (2007): 233–241. *hmg.oxfordjournals.org*. Web.

Caulfield, Mark et al. “Genome-Wide Mapping of Human Loci for Essential Hypertension.” *Lancet (London, England)* 361.9375 (2003): 2118–2123. *PubMed*. Web.

Chanda, Pritam et al. “Fast Association Tests for Genes with FAST.” *PLOS ONE* 8.7 (2013): e68585. *PLoS Journals*. Web.

Charlesworth, Deborah, and John H. Willis. “The Genetics of Inbreeding Depression.” *Nature Reviews Genetics* 10.11 (2009): 783–796. *www.nature.com*. Web.

Cheng, Alan et al. “Prospective Observational Study of Implantable Cardioverter-Defibrillators in Primary Prevention of Sudden

Cardiac Death: Study Design and Cohort Description." *Journal of the American Heart Association* 2.1 (2013): e000083. *jaha.ahajournals.org*. Web.

Chugh, Sumeet S. et al. "Epidemiology of Sudden Cardiac Death: Clinical and Research Implications." *Progress in Cardiovascular Diseases* 51.3 (2008): 213–228. *PubMed*. Web.

Coetzee, Vinet et al. "Common HLA Alleles Associated with Health, but Not with Facial Attractiveness." *PLoS ONE* 2.7 (2007): e640. *PLoS Journals*. Web.

DAWBER, T R, G F MEADORS, and F E MOORE Jr. "Epidemiological Approaches to Heart Disease: The Framingham Study." *American journal of public health and the nation's health* 41.3 (1951): 279–281. Print.

de Mutsert, Renée et al. "The Netherlands Epidemiology of Obesity (NEO) Study: Study Design and Data Collection." *European Journal of Epidemiology* 28.6 (2013): 513–523. *PubMed*. Web.

Deary, Ian J, Alan J Gow, Alison Pattie, et al. "Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936." *International journal of epidemiology* 41.6 (2012): 1576–1584. *NCBI PubMed*. Web.

Deary, Ian J, Martha C Whiteman, et al. "The Impact of Childhood Intelligence on Later Life: Following up the Scottish Mental Surveys

of 1932 and 1947." *Journal of personality and social psychology* 86.1 (2004): 130–147. *NCBI PubMed*. Web.

Deary, Ian J, Alan J Gow, Michelle D Taylor, et al. "The Lothian Birth Cohort 1936: A Study to Examine Influences on Cognitive Ageing from Age 11 to Age 70 and beyond." *BMC geriatrics* 7 (2007): 28. *NCBI PubMed*. Web.

Delaneau, Olivier et al. "Integrating Sequence and Array Data to Create an Improved 1000 Genomes Project Haplotype Reference Panel." *Nature Communications* 5 (2014): 3934. *PubMed*. Web.

Deo, Rajat, and Christine M. Albert. "Epidemiology and Genetics of Sudden Cardiac Death." *Circulation* 125.4 (2012): 620–637. *circ.ahajournals.org*. Web.

"Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group." *Controlled Clinical Trials* 19.1 (1998): 61–109. Print.

Feinleib, M et al. "The Framingham Offspring Study. Design and Preliminary Data." *Preventive medicine* 4.4 (1975): 518–525. Print.

Ferrucci, L et al. "Subsystems Contributing to the Decline in Ability to Walk: Bridging the Gap between Epidemiology and Geriatric Practice in the InCHIANTI Study." *Journal of the American Geriatrics Society* 48.12 (2000): 1618–1625. Print.

Fried, L P et al. “The Cardiovascular Health Study: Design and Rationale.” *Annals of epidemiology* 1.3 (1991): 263–276. Print.

Genes for Cerebral Hemorrhage on Anticoagulation (GOCHA) Collaborative Group. “Exploiting Common Genetic Variation to Make Anticoagulation Safer.” *Stroke; a Journal of Cerebral Circulation* 40.3 Suppl (2009): S64-66. *PubMed*. Web.

Grobbee, Diederick E. et al. “The Utrecht Health Project: Optimization of Routine Healthcare Data for Research.” *European Journal of Epidemiology* 20.3 (2005): 285–287. Print.

Grove, Megan L. et al. “Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium.” *PloS One* 8.7 (2013): e68095. *PubMed*. Web.

Harris, Tamara B et al. “Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary Applied Phenomics.” *American journal of epidemiology* 165.9 (2007): 1076–1087. *NCBI PubMed*. Web.

---. “Age, Gene/Environment Susceptibility-Reykjavik Study: Multidisciplinary Applied Phenomics.” *American journal of epidemiology* 165.9 (2007): 1076–1087. *NCBI PubMed*. Web.

Harris, Tamara B. et al. “Waist Circumference and Sagittal Diameter Reflect Total Body Fat Better Than Visceral Fat in Older Men and Women: The Health, Aging and Body Composition Study.” *Annals*

of the New York Academy of Sciences 904.1 (2000): 462–473. *Wiley Online Library*. Web.

Hendy, G. N. et al. “Mutations of the Calcium-Sensing Receptor (CASR) in Familial Hypocalciuric Hypercalcemia, Neonatal Severe Hyperparathyroidism, and Autosomal Dominant Hypocalcemia.” *Human Mutation* 16.4 (2000): 281–296. *PubMed*. Web.

Hindorff, Lucia A et al. “Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits.” *Proceedings of the National Academy of Sciences of the United States of America* 106.23 (2009): 9362–9367. *NCBI PubMed*. Web.

Hindorff, Lucia A, Elizabeth M Gillanders, and Teri A Manolio. “Genetic Architecture of Cancer and Other Complex Diseases: Lessons Learned and Future Directions.” *Carcinogenesis* 32.7 (2011): 945–954. *NCBI PubMed*. Web.

Hofman, Albert, Guy G. O. Brusselle, et al. “The Rotterdam Study: 2016 Objectives and Design Update.” *European Journal of Epidemiology* 30.8 (2015): 661–708. *PubMed*. Web.

Hofman, Albert, Monique M. B. Breteler, et al. “The Rotterdam Study: Objectives and Design Update.” *European Journal of Epidemiology* 22.11 (2007): 819–829. *PubMed Central*. Web.

Holle, R. et al. “KORA--a Research Platform for Population Based Health Research.” *Gesundheitswesen (Bundesverband Der Ärzte Des*

Öffentlichen Gesundheitsdienstes (Germany)) 67 Suppl 1 (2005): S19-25. *PubMed*. Web.

Holm, Hilma et al. "Several Common Variants Modulate Heart Rate, PR Interval and QRS Duration." *Nature Genetics* 42.2 (2010): 117–122. *PubMed*. Web.

Howie, Bryan N., Peter Donnelly, and Jonathan Marchini. "A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies." *PLOS Genet* 5.6 (2009): e1000529. *PLoS Journals*. Web.

Huang, Hailiang et al. "Gene-Based Tests of Association." *PLoS genetics* 7.7 (2011): e1002177. *PubMed*. Web.

John, U et al. "Study of Health In Pomerania (SHIP): A Health Examination Survey in an East German Region: Objectives and Design." *Sozial- und Präventivmedizin* 46.3 (2001): 186–194. Print.

Jørgensen, Torben et al. "A Randomized Non-Pharmacological Intervention Study for Prevention of Ischaemic Heart Disease: Baseline Results Inter99." *European Journal of Cardiovascular Prevention and Rehabilitation: Official Journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology* 10.5 (2003): 377–386. *PubMed*. Web.

Juster, F. Thomas, and Richard Suzman. “An Overview of the Health and Retirement Study.” *The Journal of Human Resources* 30 (1995): S7–S56. *JSTOR*. Web.

Kannel, W B et al. “An Investigation of Coronary Heart Disease in Families. The Framingham Offspring Study.” *American journal of epidemiology* 110.3 (1979): 281–290. Print.

Kapoor, Ashish et al. “An Enhancer Polymorphism at the Cardiomyocyte Intercalated Disc Protein NOS1AP Locus Is a Major Regulator of the QT Interval.” *American Journal of Human Genetics* 94.6 (2014): 854–869. *PubMed*. Web.

Kim, Jong Wook et al. “A Common Variant in SLC8A1 Is Associated with the Duration of the Electrocardiographic QT Interval.” *American Journal of Human Genetics* 91.1 (2012): 180–184. *PubMed*. Web.

Kopito, R. R. et al. “Regulation of Intracellular pH by a Neuronal Homolog of the Erythrocyte Anion Exchanger.” *Cell* 59.5 (1989): 927–937. Print.

Lie, Hanne C., Leigh W. Simmons, and Gillian Rhodes. “Does Genetic Diversity Predict Health in Humans?” *PLoS ONE* 4.7 (2009): e6391. *PLoS Journals*. Web.

Luo, G., J. Q. Zhang, et al. “Complete cDNA Sequence and Tissue Localization of N-RAP, a Novel Nebulin-Related Protein of Striated

Muscle." *Cell Motility and the Cytoskeleton* 38.1 (1997): 75–90. *PubMed*. Web.

Luo, G., E. Leroy, et al. "Mapping of the Gene (NRAP) Encoding N-RAP in the Mouse and Human Genomes." *Genomics* 45.1 (1997): 229–232. *PubMed*. Web.

Menard, Scott. "Six Approaches to Calculating Standardized Logistic Regression Coefficients." *The American Statistician* 58.3 (2004): 218–223. *CrossRef*. Web.

Mitton, J. B., and M. C. Grant. "Associations Among Protein Heterozygosity, Growth Rate, and Developmental Homeostasis." *Annual Review of Ecology and Systematics* 15 (1984): 479–499. Print.

Moayyeri, Alireza et al. "The UK Adult Twin Registry (TwinsUK Resource)." *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies* 16.1 (2013): 144–149. *PubMed*. Web.

Mond, Harry G., and Alessandro Proclemer. "The 11th World Survey of Cardiac Pacing and Implantable Cardioverter-Defibrillators: Calendar Year 2009--a World Society of Arrhythmia's Project." *Pacing and clinical electrophysiology: PACE* 34.8 (2011): 1013–1027. *PubMed*. Web.

Mount, D. B. et al. “Cloning and Characterization of KCC3 and KCC4, New Members of the Cation-Chloride Cotransporter Gene Family.” *The Journal of Biological Chemistry* 274.23 (1999): 16355–16362. Print.

Newton-Cheh, Christopher, Mark Eijgelsheim, et al. “Common Variants at Ten Loci Influence QT Interval Duration in the QTGEN Study.” *Nature Genetics* 41.4 (2009): 399–406. *PubMed*. Web.

Newton-Cheh, Christopher, Martin G. Larson, et al. “QT Interval Is a Heritable Quantitative Trait with Evidence of Linkage to Chromosome 3 in a Genome-Wide Linkage Analysis: The Framingham Heart Study.” *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 2.3 (2005): 277–284. *PubMed*. Web.

Nishida, T. et al. “Characterization of a Novel Mammalian SUMO-1/Smt3-Specific Isopeptidase, a Homologue of Rat Axam, Which Is an Axin-Binding Protein Promoting Beta-Catenin Degradation.” *The Journal of Biological Chemistry* 276.42 (2001): 39060–39066. *PubMed*. Web.

Nolte, Ilja M. et al. “Common Genetic Variation near the Phospholamban Gene Is Associated with Cardiac Repolarisation: Meta-Analysis of Three Genome-Wide Association Studies.” *PloS One* 4.7 (2009): e6138. *PubMed*. Web.

Noseworthy, Peter A. et al. “Common Genetic Variants, QT Interval, and Sudden Cardiac Death in a Finnish Population-Based Study.” *Circulation. Cardiovascular Genetics* 4.3 (2011): 305–311. *PubMed*. Web.

“Online Mendelian Inheritance in Man.” N.p., n.d. Web. 17 Jan. 2014.

Pardo, L. M. et al. “The Effect of Genetic Drift in a Young Genetically Isolated Population.” *Annals of Human Genetics* 69.Pt 3 (2005): 288–295. *PubMed*. Web.

Park, Chong Yon et al. “skNAC, a Smyd1-Interacting Transcription Factor, Is Involved in Cardiac Development and Skeletal Muscle Growth and Regeneration.” *Proceedings of the National Academy of Sciences of the United States of America* 107.48 (2010): 20750–20755. *PubMed*. Web.

Pers, Tune H. et al. “Biological Interpretation of Genome-Wide Association Studies Using Predicted Gene Functions.” *Nature Communications* 6 (2015): 5890. *www.nature.com*. Web.

Pfeufer, Arne et al. “Common Variants at Ten Loci Modulate the QT Interval Duration in the QTSCD Study.” *Nature Genetics* 41.4 (2009): 407–414. *PubMed*. Web.

Piertney, S. B., and M. K. Oliver. “The Evolutionary Ecology of the Major Histocompatibility Complex.” *Heredity* 96.1 (2005): 7–21. *www.nature.com*. Web.

Porta, Alberto et al. “Autonomic Control of Heart Rate and QT Interval Variability Influences Arrhythmic Risk in Long QT Syndrome Type 1.” *Journal of the American College of Cardiology* 65.4 (2015): 367–374. *Silverchair*. Web.

Psaty, Bruce M. et al. “Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts.” *Circulation: Cardiovascular Genetics* 2.1 (2009): 73–80. *circgenetics.ahajournals.org*. Web.

Purcell, Shaun et al. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *The American Journal of Human Genetics* 81.3 (2007): 559–575. *ScienceDirect*. Web.

Raitakari, Olli T. et al. “Cohort Profile: The Cardiovascular Risk in Young Finns Study.” *International Journal of Epidemiology* 37.6 (2008): 1220–1226. *PubMed*. Web.

Roberts, S. Craig et al. “MHC-Heterozygosity and Human Facial Attractiveness.” *Evolution and Human Behavior* 26.3 (2005): 213–226. *ScienceDirect*. Web.

Scholtens, Salome et al. “Cohort Profile: LifeLines, a Three-Generation Cohort Study and Biobank.” *International Journal of Epidemiology* 44.4 (2015): 1172–1180. *PubMed*. Web.

Schwartz, Peter J., Lia Crotti, and Roberto Insolia. “Long-QT Syndrome: From Genetics to Management.” *Circulation. Arrhythmia and Electrophysiology* 5.4 (2012): 868–877. *PubMed*. Web.

Smith, Blair H. et al. “Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The Study, Its Participants and Their Potential for Genetic Research on Health and Illness.” *International Journal of Epidemiology* 42.3 (2013): 689–700. *PubMed*. Web.

Sotoodehnia, Nona et al. “Common Variants in 22 Loci Are Associated with QRS Duration and Cardiac Ventricular Conduction.” *Nature Genetics* 42.12 (2010): 1068–1076. *PubMed*. Web.

Splansky, Greta Lee et al. “The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: Design, Recruitment, and Initial Examination.” *American journal of epidemiology* 165.11 (2007): 1328–1335. *NCBI PubMed*. Web.

Stevenson, Lynne Warner. “Implantable Cardioverter-Defibrillators for Primary Prevention of Sudden Death in Heart Failure.” *Circulation* 114.2 (2006): 101–103. *circ.ahajournals.org*. Web.

Stouffer Samuel, A. et al. *The American Soldier. Adjusting During Army Life, Vol. 1*. Princeton, Princeton University Press, 1949. Print.

Szulkin, Marta, Nicolas Bierne, and Patrice David. “Heterozygosity-Fitness Correlations: A Time for Reappraisal.” *Evolution* 64.5 (2010): 1202–1217. *Wiley Online Library*. Web.

Takata, Hajime et al. “INFLUENCE OF MAJOR HISTOCOMPATIBILITY COMPLEX REGION GENES ON HUMAN LONGEVITY AMONG OKINAWAN-JAPANESE CENTENARIANS AND NONAGENARIANS.” *The Lancet* 330.8563 (1987): 824–826. *ScienceDirect*. Web.

Taylor, Herman A. et al. “Toward Resolution of Cardiovascular Health Disparities in African Americans: Design and Methods of the Jackson Heart Study.” *Ethnicity & Disease* 15.4 Suppl 6 (2005): S6-4–17. Print.

“The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives. The ARIC Investigators.” *American journal of epidemiology* 129.4 (1989): 687–702. Print.

Tobin, Martin D. et al. “Common Variants in Genes Underlying Monogenic Hypertension and Hypotension and Blood Pressure in the General Population.” *Hypertension* 51.6 (2008): 1658–1664. *PubMed*. Web.

Voight, Benjamin F. et al. “The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits.” *PLoS genetics* 8.8 (2012): e1002793. *PubMed*. Web.

Völzke, Henry et al. “Cohort Profile: The Study of Health in Pomerania.” *International journal of epidemiology* 40.2 (2011): 294–307. *NCBI PubMed*. Web.

Wichmann, H.-E. et al. “KORA-Gen--Resource for Population Genetics, Controls and a Broad Spectrum of Disease Phenotypes.” *Gesundheitswesen (Bundesverband Der Ärzte Des Öffentlichen Gesundheitsdienstes (Germany))* 67 Suppl 1 (2005): S26-30. *PubMed*. Web.

Zemunik, Tatijana et al. “Genome-Wide Association Study of Biochemical Traits in Korcula Island, Croatia.” *Croatian Medical Journal* 50.1 (2009): 23–33. Print.

Zhang, Hong, Hisato Saitoh, and Michael J. Matunis. “Enzymes of the SUMO Modification Pathway Localize to Filaments of the Nuclear Pore Complex.” *Molecular and Cellular Biology* 22.18 (2002): 6498–6508. Print.

Curriculum vitae

Nathan A Bihlmeyer
Miller Research Building, Room 420 ● 733 N Broadway, Baltimore, MD 21205
Lab: (443) 287-4434 ● https://nathan.a.bihlmeyer.me/ ● nbihlme1@jhmi.edu

Education

Johns Hopkins University School of Medicine, Baltimore, MD
McKusick-Nathans Institute of Genetic Medicine
Doctor of Philosophy in Human Genetics and Molecular Biology, 2011 – 2016
National Science Foundation (NSF) Graduate Research Fellow

North Carolina State University, Raleigh, NC
Bachelor of Science in Biochemistry, Minor in Genetics, 2007 – 2011
Summa Cum Laude, Barry M. Goldwater Scholar

Research Experience

Graduate Student, 8/2011 – 10/2016
Laboratory of Dan E Arking, PhD
McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD
- Elucidating genetic determinants of heart disease and longevity through genome-wide association analyses in humans
- Study the genetics of electrical conduction, focusing on the role both rare and common genetic variants
- Run projects within the CHARGE consortium EKG and Aging & Longevity working groups, supporting up to 23 other analysts

Rotation Student, 8/2011 – 12/2011
Laboratory of Nicholas Marsh-Armstrong, PhD
Hugo W Moser Research Institute, Kennedy Krieger Institute, Baltimore, MD
- Investigate glial cell degradation of nerve axon mitochondria in optic nerve from *Mus musculus* and *Xenopus*
- Optimize biochemical assays

Undergraduate Student, 5/2008 – 7/2011
Laboratory of Robert G Franks, PhD
Department of Genetics, North Carolina State University, Raleigh, NC
- Determine how the SEUSS family of transcriptional regulators affects the carpel margin meristem, the meristem that gives rise to seeds, in *Arabidopsis thaliana*
- Perform biochemical assays with bacterially expressed fusion proteins
- Independent research in gene regulation and design of experiment protocols

Skills

Notable Software and Programming Languages

R	Perl	Python
Java	Bash	IGV
Plink	Impute2	ShapeIT
DEPICT	Eigenstrat	Epacts
FAST	R-seqMeta	R-survival
METAL	RAREMETAL	OpenGridEngine
Affymetrix Power Tools	NCBI SRA Toolkit	

Notable Molecular Biology Techniques

Dissection of optic nerve	Cryosectioning
in situ Hybridization	FISH with Mito DNA
TUNEL assay	Immunohistochemistry
Protein *in vitro* expression	Protein-protein interaction assays
Protein Solubilization	Phenotypic Analysis
Western Blotting	Affinity Chromatography
Sub-cloning	*E. coli* Culturing and Induction
Yeast 2-Hybrid	Oligo Design
Confocal Microscopy	PCR Amplification

Bimolecular Fluorescence Complementation (BiFC)

Publications

Lubitz, Steven A. et al. "Whole Exome Sequencing in Atrial Fibrillation." *PLOS Genet* 12.9 (2016): e1006284.

Busby, Ben et al. "Closing Gaps between Open Software and Public Data in a Hackathon Setting: User-Centered Software Prototyping." *F1000Research* 5 (2016): 672.

Wessel, Jennifer et al. "Low-Frequency and Rare Exome Chip Variants Associate with Fasting Glucose and Type 2 Diabetes Susceptibility." *Nature Communications* 6 (2015): 5897.

Bihlmeyer, Nathan A. et al. "Genetic Diversity Is a Predictor of Mortality in Humans." *BMC genetics* 15 (2014): 159. ◎

Davis, Chung-ha O. et al. "Transcellular Degradation of Axonal Mitochondria." *Proceedings of the National Academy of Sciences of the United States of America* 111.26 (2014): 9633–9638.

In preparation

Li, Man et al. “Two loci identified through large-scale exome chip analysis, *SOS2* and *ACP1*, show evidence for altered kidney development and function.” *Journal of American Society of Nephrology* (Accepted).

Christophersen, Ingrid E. et al. “Common and Rare Variant Analyses in >150,000 Individuals Reveal Twelve Novel Genetic Loci Associated with Atrial Fibrillation.” *Nature Genetics* (Under Review).

Bihlmeyer, Nathan A. et al. “Exome-wide analysis of 95,626 individuals identifies 10 novel loci associated with QT and JT intervals.” *PLoS Genetics* (Drafted). ◎

Prins, Bram P. et al. “Novel association of rare variants in *ADAMTS6* with cardiac conduction.” *Nature Genetics* (Drafted).

Conferences

Oral Presentations

Bihlmeyer, Nathan A. et al. “Investigating the effects of coding variants on QT and JT intervals utilizing data from 95,626 individuals” *American Society of Human Genetics Annual Meeting* (2015). ◎

Bihlmeyer, Nathan A. et al. “Characterization of Physical Interactions within a Multimeric Transcriptional Corepressor Complex in *Arabidopsis*” *Fourth Annual Atlantic Coast Conference Meeting of the Minds Conference* (2009).

Poster Presentations

Bihlmeyer, Nathan A. et al. “Genetic Diversity is a Predictor of Survival in Humans.” *American Society of Human Genetics Annual Meeting* (2013). ◎

Bihlmeyer, Nathan A. et al. “Whole Genome Heterozygosity as a Predictor of Longevity.” *6th Annual Symposium and Poster Session on Genomics and Bioinformatics by the Center for Computational Genomics at Johns Hopkins* (2012).
Presentation was awarded “Best poster in the category of biology.”

Bihlmeyer, Nathan A. et al. “Exploration of Transcriptional Control Pathways Involving SEUSS in *Arabidopsis thaliana*.” *State of North Carolina Undergraduate Research and Creativity Symposium* (2010).

Bihlmeyer, Nathan A. et al. “Bimolecular Fluorescence Complementation as a Method to Study Transcriptional Regulatory Complexes.” *North Carolina State University’s Annual Genetics Department Retreat* (2010).
Presentation was awarded “Potential to be Great Award.”

Involvement

2015

- American Society of Human Genetics (ASHG) Annual Meeting
- NCBI Hackathon ◎

2014

- Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Investigator Meeting: DC, November
- Genomics of Common Diseases (GCD) Conference
- Teaching assistant of Introduction to Computational Genetics course ◎
- Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Investigator Meeting: LA, January

2013

- American Society of Human Genetics (ASHG) Annual Meeting

2012

- 6th Annual Symposium and Poster Session on Genomics and Bioinformatics by the Center for Computational Genomics at Johns Hopkins
- Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Analysis Workshop: Boston, September
- 53rd Annual Short Course on Medical and Experimental Mammalian Genetics being held at The Jackson Laboratory in Bar Harbor Maine

Undergraduate Activities and Societies

- President of the Genetics Club
- Vice President of the Biochemistry Club
- Phi Beta Kappa Honors Society
- National Society of Collegiate Scholars
- College of Agriculture & Life Sciences Honors Program
- Biochemistry Honors Program
- North Carolina State University Presidents' Round Table

Honors

2011

- National Science Foundation Graduate Research Fellowship
- The H. Robert Horton Award for Outstanding Undergraduate Achievement in Biochemistry

2010

- Barry M Goldwater Scholarship

2009

- Undergraduate Research Award from North Carolina State University Office of Undergraduate Research (Grant for research)

2006

- Eagle Scout from the Boy Scouts of America

References

Dan E Arking, PhD
Associate Professor
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine
(410) 502-4867
arking@jhmi.edu

Nona Sotoodehnia, MD MPH
Laughlin Endowed Professor in Cardiology
Co-Director, Cardiovascular Health Research Unit
Harborview Medical Center
University of Washington
(206) 221-7775
nsotoo@u.washington.edu

David Valle, MD
Henry J. Knott Professor
Director of the McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine
(410) 955-4260
dvalle@jhmi.edu

Reprint Authorization

Appendix A: (Bihlmeyer et al.)